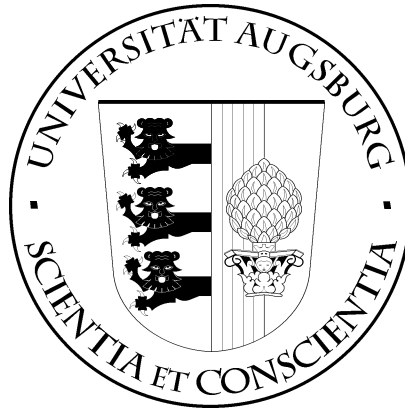


UNIVERSITÄT AUGSBURG



**Vom physiologischen Signal zur Emotion:
Implementierung und Vergleich ausgewählter
Methoden zur Merkmalsextraktion und Klassifikation**

Johannes Wagner

Report 2005-1

März 2005



INSTITUT FÜR INFORMATIK

D-86135 AUGSBURG

BACHELORARBEIT

**Vom physiologischen Signal zur Emotion:
Implementierung und Vergleich ausgewählter
Methoden zur Merkmalsextraktion und Klassifikation**

vorgelegt von

Johannes Wagner
(Matr.Nr.: 731823)

zur Erlangung des akademischen Grades
- Bachelor of Science (B.Sc.) -
an der Fakultät für Angewandte Informatik
der Universität Augsburg

Betreuer: Dr. Jong-Hwa Kim

Augsburg, 28. 11. 2004

ZUSAMMENFASSUNG

Vom physiologischen Signal zur Emotion: Implementierung und Vergleich ausgewählter Methoden zur Merkmalsextraktion und Klassifikation

Dem Computer die Fähigkeit zu verleihen, die Emotionen seines Benutzers zu erkennen, stellt eine große Herausforderung dar, die seit vielen Jahren die Forschung auf diesem Gebiet vorantreibt. Bisher wurden vor allem die Gesicht- und Spracherkennung fokussiert. Emotionen aus physiologischen Daten abzuleiten, ist erst in den letzten Jahren als eine weitere Komponente dazugestoßen. Dabei gibt es einige Vorteile gegenüber den herkömmlichen Methoden. Da die biologischen Reaktionen unseres Körpers durch das zentrale Nervensystem gesteuert werden, sind sie nur teilweise oder gar nicht der Kontrolle unseres Bewusstseins unterlegen. So besteht eine direkte Verbindung zu unseren Emotionen, die einen unmittelbaren Einblick in unseren affektiven Zustand gewährt. Die vorliegende Arbeit erörtert alle Schritte, um ein System zu entwickeln, das in der Lage ist, emotionsspezifische Muster aus physiologischen Signalen zu extrahieren und zu klassifizieren. Dabei werden jeweils verschiedene bekannte Verfahren aus der Mustererkennung getestet und miteinander verglichen. Die Evaluation der Implementierung erfolgt an einem Korpus, der physiologische Daten zu den vier affektiven Zuständen Freude, Ärger, Trauer und Glückseligkeit enthält. Für die Aufzeichnung des Korpus wurden Signale von den vier Sensoren EKG (Elektrokardiogramm), EMG (Elektromyogramm), RSP (Atmung) und SC (Hautleitwert) gesammelt. Dabei wurde auf eine musikalische Induzierung zur Auslösung der Emotionen zurückgegriffen.

Inhaltsverzeichnis

1	Einleitung	7
1.1	Emotionserkennung aus physiologischen Daten	8
1.2	Vom physiologischen Signal zur Emotion	8
2	Menschliche Emotionen	10
2.1	Kognitiv-physiologische Theorie	10
2.2	Funktionen und Auswirkungen	11
2.3	Emotionsmodelle	12
3	Affective Computing	14
3.1	Neue Möglichkeiten der Interaktion	14
3.2	Forschung und Anwendung	15
3.2.1	Tragbare Systeme	15
3.2.2	Emotionserkennung	16
3.2.3	Affektive Anwendungen	18
4	Emotionserkennung mit physiologischen Daten	22
4.1	Motivation	22
4.1.1	Verwandte Arbeiten	23
4.1.2	Schwierigkeiten und Grenzen	23
4.1.3	Vorteile gegenüber herkömmlichen Verfahren	24
4.2	Sensoren und Signale	24
4.2.1	Hautleitwert (SC)	25
4.2.2	Muskelkontraktion (EMG)	27
4.2.3	Atmung (RSP)	29
4.2.4	Herzschlag (EKG)	30
5	Aufbau des Datenkorpus	33
5.1	Vorüberlegungen zum Korpus	33
5.1.1	Auswahl	34
5.1.2	Aufnahme	34

5.1.3	Erregung	36
5.1.4	Reihenfolge	37
5.2	Beschreibung des Experiments	37
5.3	Beschreibung des Korpus	38
6	Vom Signal zur Emotion	40
6.1	Berechnung relevanter Features	40
6.1.1	Basisfeatures	41
6.1.2	Berechnung der Atemfrequenz	43
6.1.3	Berechnung der Herzfrequenz	43
6.1.4	Zusätzliche Features	43
6.1.5	Grafische Analyse der Features	47
6.2	Reduktion des Feature-Raums	48
6.2.1	Überblick über gängige Methoden	49
6.2.2	In dieser Arbeit eingesetzte Verfahren	50
6.3	Mustererkennung	56
6.3.1	Überblick über gängige Methoden	57
6.3.2	In dieser Arbeit eingesetzte Verfahren	59
7	Ergebnisse	66
7.1	Test von Klassifikationsverfahren anhand des eigenen Korpus	66
7.1.1	Vier Emotionen ohne Dimensionsreduktion	66
7.1.2	Vier Emotionen mit Dimensionsreduktion	68
7.1.3	Emotionsgruppen mit Dimensionsreduktion	69
7.2	Bewertung der Features	69
7.3	Vergleich mit dem Korpus des MIT	71
8	Schluss	74

Abbildungsverzeichnis

2.1	Die Achsen des 2- und 3-dimensionalen Emotionsmodelles.	13
3.1	Beispiele für tragbare Sensoren. Quelle: [12]	16
3.2	Gesichtserkennung durch Verfolgung von Referenzpunkten (Quelle: [13]).	17
3.3	Feature-Extraktion aus einem Sprachsignal. Die Abbildung wurde erstellt mit dem Programm Praat. http://www.fon.hum.uva.nl/praat/	17
3.4	Physiologische Signale in den Emotionen Freude und Trauer.	18
3.5	Abbildung eines affektiven Desktops. Quelle: [15]	19
3.6	Eine affektive Anwendung: der „Affective DJ”. Quelle: [17]	20
3.7	Eine affektive Anwendung: der „Galvactivator”. Quelle: [18]	21
4.1	Der Verlauf eines typischen SC-Signals.	25
4.2	Anbringung der Elektroden zur Aufzeichnung des SC-Signals.	26
4.3	Der Verlauf eines typischen EMG-Signals.	27
4.4	Anbringung der Elektroden zur Aufzeichnung des EMG-Signals.	28
4.5	Der Verlauf eines typischen RSP-Signals.	29
4.6	Anbringung des Bauchgurts zur Aufzeichnung des RSP-Signals.	30
4.7	Der Verlauf eines typischen EKG-Signals.	31
4.8	Anbringung der Elektroden zur Aufzeichnung des EKG-Signals.	31
5.1	Das Emotionsmodell für die Aufzeichnung des eigenen Korpus.	34
5.2	Schematischer Ablauf des Experiments zur Aufzeichnung des eigenen Korpus.	38
6.1	Schematischer Ablauf der Mustererkennung.	41
6.2	Berechnung der Atemfrequenz.	42
6.3	EKG mit Power-Spektrum	43
6.4	Berechnung der SC-Grundlinie.	44
6.5	Boxplot von Mittelwert und Varianz des SC-Signals.	45
6.6	Entfernung von Artefakten aus dem EMG-Signal.	46

6.7	Berechnung der Amplitude aus dem RSP-Signal.	47
6.8	Boxplot der statistischen Verteilung der mittleren Atemfrequenz zwischen den Klassen.	48
6.9	Verlauf des Klassifikationsfehler mit ANOVA in Abhängigkeit von der Featurezahl.	52
6.10	Verlauf des Klassifikationsfehler mit SFS und SBS in Abhängigkeit von der Featurezahl.	52
6.11	Verlauf des Klassifikationsfehlers nach Reduktion des Feature- Raums mit PCA.	53
6.12	Dimensionsreduktion 2-dimensionaler Punktwolken.	55
6.13	Verlauf des Klassifikationsfehlers nach Dimensionsreduktion mit der Feature-Projektion.	56
6.14	Schematischer Ablauf der Klassifikation.	57
6.15	Aufbau eines mehrschichtigen Perzeptrons.	60
6.16	Repräsentation des Klassifikators durch eine Menge von Diskrimi- nanzfunktionen.	61
6.17	Klassifikation mit einem Wahrscheinlichkeitsmodell.	62
6.18	Klassifikation mit kNN.	63
6.19	Netzausgabe eines mehrschichtigen Perzeptrons.	65

Tabellenverzeichnis

2.1	Basisemotionen verschiedener Wissenschaftler.	12
7.1	Übersicht über die Ergebnisse verwandter Arbeiten.	67
7.2	Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen ohne Dimensionsreduktion.	68
7.3	Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen mit Dimensionsreduktion.	69
7.4	Vergleich der Erkennungsquoten für verschiedene Klassifikatoren und Emotionsgruppen mit Dimensionsreduktion.	69
7.5	Die Tabelle zeigt für jede Emotion die Features, die entsprechend dem Test von Tukey bei einem Signifikanzlevel von 0,05, als signifikant eingestuft wurden.	70
7.6	Vergleich zwischen MIT- und AuDB-Korpus.	71
7.7	Überblick über die Emotionen aus dem Datenkorpus des MIT (vgl. Tabelle 3.1 in [19]).	72
7.8	Vergleich der Erkennungsquoten für verschiedene Klassifikatoren und Emotionsgruppen mit Dimensionsreduktion.	72
7.9	Die Tabelle zeigt für die Ergebnisse des Tukey-Signifikanztest für den MIT- und den eigenen Korpus bei einem Signifikanzlevel von 0,05 für hohe Erregung im Vergleich zu schwacher bzw. positive gegenüber negative Valenz.	73
8.1	Klassifikation von vier Emotionen mit LDF ohne Dimensionsreduktion.	75
8.2	Klassifikation von vier Emotionen mit 3NN ohne Dimensionsreduktion.	75
8.3	Klassifikation von vier Emotionen mit 5NN ohne Dimensionsreduktion.	75
8.4	Klassifikation von vier Emotionen mit 10NN ohne Dimensionsreduktion.	76
8.5	Klassifikation von vier Emotionen mit 15NN ohne Dimensionsreduktion.	76

8.6	Klassifikation von vier Emotionen mit MLP2 ohne Dimensionsreduktion.	76
8.7	Klassifikation von vier Emotionen mit MLP4 ohne Dimensionsreduktion.	76
8.8	Klassifikation von vier Emotionen mit MLP6 ohne Dimensionsreduktion.	77
8.9	Klassifikation von vier Emotionen mit MLP8 ohne Dimensionsreduktion.	77
8.10	Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen mit Dimensionsreduktion in Abhängigkeit der selektierten Features.	78

Kapitel 1

Einleitung

It is time to examine how emotions can be incorporated into models of intelligence, and particularly, into computers and their interactions with humans.

Rosalind W. Picard, von der dieses Zitat aus dem Jahre 1997 stammt, war eine der ersten Forscher, die sich Gedanken darüber machte, wie sich Emotionen in Zukunft nutzen lassen, um die Interaktion von Mensch und Maschine zu verändern [22]. Ihrer Vorstellung nach sollen intelligente Agenten in Zukunft den Computer nicht mehr als kaltes und gefühlloses Arbeitsgerät, sondern als persönlichen Assistenten und Vertrauten, der sich um unsere Bedürfnisse und Wünsche kümmert, erscheinen lassen. Tragbare Sensoren, integriert in unsere Kleidung, werden uns in der Arbeit, beim Einkaufen oder in unserer Freizeit begleiten und Mikrosystemen Informationen über unser Befinden und Gefühle liefern. Zusammen mit anderer Kontextinformation sollen diese mobilen Begleiter lernen auf unsere Verhaltensmuster und Wünsche einzugehen, in unserem Sinne lästige Routinearbeit verrichten und uns bei der Optimierung täglicher Arbeitsschritte helfen. Wir stehen damit vor dem Beginn einer grundlegenden Wandlung in der Interaktion zwischen Mensch und Maschine. Diese ist unumgänglich, denn in Zukunft werden noch mehr Menschen, noch mehr Zeit mit dem Computer verbringen. Doch viele Menschen haben Probleme mit der logischen und rationalen Art, die die Arbeit mit dem Computer verlangt. Es ist also notwendig die Grundlage für eine natürlichere Kommunikation zwischen Mensch und Maschine zu schaffen. Dafür muss der Computer jedoch zunächst lernen, auf unsere Gefühle und Empfindungen zu reagieren. Ermöglicht wird dies durch neue Systeme, die den affektiven Zustand des Benutzers jederzeit schnell und zuverlässig erfassen können. Die klassischen Methode zur Emotionserkennung sind die Gesichts- und Spracherkennung. Emotionen aus physiologischen Daten abzuleiten, ist erst in den letzten Jahren als weitere Komponente dazugestoßen.

1.1 Emotionserkennung aus physiologischen Daten

Physiologische Daten zur Emotionserkennung zu verwenden ist erst seit wenigen Jahren ein Thema in der Forschung. Dabei gibt es einige Vorteile gegenüber den herkömmlichen Methoden. Da die biologischen Reaktionen unseres Körpers durch das zentrale Nervensystem gesteuert werden, sind sie nur teilweise oder gar nicht der Kontrolle unseres Bewusstseins unterlegen. Es besteht also eine direkte und unverfälschte Verbindung zu unseren Emotionen, die einen direkten Einblick in unseren affektiven Zustand gewährt. Auch außerhalb der Reichweite von Kameras und Mikrofonen können mobile Sensoren, die direkt in Kleidung oder Schmuck integriert sind, einen kontinuierlichen Signalstrom garantieren. So kann die Aufzeichnung unserer physiologischen Daten in jeder beliebigen Umgebung fortgeführt werden: z. B. in der Arbeit, beim Einkaufen oder während der Autofahrt. Noch herrscht unter den Psychologen Streit darüber, ob sich die physiologischen Reaktionen des Menschen eindeutig mit bestimmten Emotionen verknüpfen lassen. Da der Mensch sehr individuell auf Emotionen reagieren kann, ist es schwer allgemeingültige Regeln zu finden. Ein adaptives System, das sich seinem jeweiligen Benutzer anpasst und mit der Zeit ein persönliches Bild von ihm entwirft, hat daher wesentlich bessere Aussichten auf Erfolg. Erste Untersuchungen von physiologischen Daten mit Hilfe digitaler Signalverarbeitungstechniken in Kombination mit Mustererkennungsalgorithmen liefern auch Erfolg versprechende Ergebnisse. An dieser Stelle sei auf die Arbeiten von Healey [19] und Vyzas [20] oder Lisetti und Nasoz [30] verwiesen.

1.2 Vom physiologischen Signal zur Emotion

Diese Arbeit beschäftigt sich mit allen Schritten, die zum Entwurf eines Systems, das Emotionen aus physiologischen Daten erkennen kann, nötig sind. Die ersten beiden Kapitel bieten dabei zunächst einen kurzen Überblick über menschliche Emotionen und welche Rolle sie in dem neuen Forschungsgebiet des Affective Computing spielen. Dabei wird insbesondere darauf eingegangen, welche neuen Möglichkeiten sich für die Mensch-Maschine-Kommunikation ergeben und wie der momentane Stand der Forschung aussieht. Anschließend folgt eine Einführung in die Emotionserkennung mit physiologischen Daten und eine ausführliche Beschreibung der Sensoren und Signale. Eine Aufgabe dieser Arbeit bestand in der Aufzeichnung eines eigenen Datenkorpus mit physiologischen Daten. Im nächsten Kapitel werden deshalb wichtige Punkte, die bei der Erstellung des Korpus berücksichtigt wurden, besprochen. Aufbauend auf den Datenkorpus werden anschließend Methoden der Signalverarbeitung zur Auswertung der aufgezeichneten Daten beschrieben. Insgesamt 32 Features wurden berechnet und anschließend mit verschiedenen Techniken

der Dimensionsreduktion für den Einsatz zum Training eines Klassifikators vorbereitet. Zur Klassifikation wurden sowohl statistische Verfahren, als auch neuronale Netze eingesetzt. Das letzte Kapitel stellt die Ergebnisse dar, die mit den implementierten Techniken erreicht wurden. Die Erfolgsquote lag dabei für vier Emotionen bei über 90%. Für bestimmte Emotionsgruppen konnte eine Rate von über 95% erzielt werden. Mit Methoden der Statistik werden schließlich noch Zusammenhänge zwischen Emotionen und bestimmten signifikanten Änderungen in den physiologischen Daten untersucht. Zur Festigung der erzielten Ergebnisse wird außerdem noch ein weiterer Datenkorpus, der am MIT aufgezeichnet wurde, ausgewertet und mit den Ergebnissen des eigenen Korpus verglichen.

Kapitel 2

Menschliche Emotionen

Lange wurde die Rolle, die Emotionen in unserem Leben spielen, unterschätzt. Mittlerweile sind sich Psychologen einig: Emotionen haben entscheidenden Einfluss auf die Kommunikation mit unseren Mitmenschen, helfen uns Entscheidungen zu treffen, verändern unsere Leistungsfähigkeit und beeinflussen Lern- und Denkprozesse. Seit Jahrhunderten beschäftigen sich Wissenschaftler mit der Frage wie Emotionen entstehen und welche Funktion sie in unser Leben übernehmen. Doch nach wie vor gibt es wenige verlässliche Antworten. Tatsächlich könnte die Uneinigkeit unter den Wissenschaftlern kaum größer sein. P.R. Kleinginna und A.M. Kleinginna zählten 1981 in der Fachliteratur 92 verschiedene Emotionsdefinitionen, die sie in immerhin noch zehn unversöhnliche Kategorien einordneten [1]. Im Folgenden wird eine, bis heute populäre Definition, die auf William James zurückgeht, vorgestellt. Weiter beschäftigt sich dieses Kapitel mit den Funktionen und Auswirkungen von Emotionen und gibt einen Überblick über die gebräuchlichsten Emotionsmodelle.

2.1 Kognitiv-physiologische Theorie

Die kognitiv-physiologische Theorie geht auf den Philosophen und Psychologen William James (1884) zurück. In seinem Modell geht er davon aus, dass die Wahrnehmung einer erregenden Tatsache eine körperliche Veränderung in unserem Organismus hervorruft. Weiter glaubt er, dass diese Veränderungen emotionsspezifisch sind und wir das bewusste Erleben der körperlichen Veränderungen als Emotion erfahren. Cannon, ein Schüler von James, beobachtete jedoch, dass mit verschiedenen emotionalen und nicht-emotionalen Zuständen dieselben körperlichen Veränderungen einhergehen (z. B. Zittern vor Angst, aber auch vor Wut). Stanley Schachter vermutet deshalb in seiner Theorie (1964), dass auch die physiologische Erregung im peripheren Nervensystem und deren Empfindung eine Rolle spielen. Ob und

welche Emotion erzeugt wird, hängt seiner Vorstellung nach allein von der Kognition ab. Die Kognition befähigt uns zu einer Bewertung der erlebten Situation und führt dann bei gleichzeitigem Vorhandensein einer Errungsempfindung zu der erlebten Emotion. In der „Metapher des Musikautomaten“ versucht Mandler diesen Zusammenhang zu veranschaulichen. Der Einwurf einer Münze in den Automaten entspricht der Erregung, die Wahl jedoch, welche Platte gespielt wird, ist unabhängig davon. Diese entspricht gerade der Kognition über die auslösende Situation. Die kognitiv-physiologischen Theorien sind nur eine Teilmenge vieler Theorien, die versuchen die Entstehung von Emotionen zu erklären. Für viele der Annahmen gibt es bis heute keine beweisenden Studien, aber auch keine, die sie widerlegen. Die Auffassung, dass Emotionen primär auf Kognitionen beruhen, ist jedoch bis heute ein feste Säule in der Emotionspsychologie geblieben. Für weitere Informationen zu den kognitiv-physiologischen Theorien und anderen Emotionsmodellen sei auf [2] verwiesen.

2.2 Funktionen und Auswirkungen

Natürlich stellt sich für die Emotionsforscher auch die Frage, welchen Sinn Emotionen in unserem Leben erfüllen. Man kann Emotionen als eine Art Verhaltensprogramm sehen, das uns eine bedürfnis- und situationsgerechte Auswahl von Verhaltensweisen und die Regulierung von deren Intensität und Ausdauer ermöglicht [2]. Sie übernehmen damit eine ähnliche Funktion wie unsere Instinkte. Welche Auswirkungen Emotionen auf unser Verhalten haben, kann jeder an sich selbst beobachten, z. B. wenn unsere Konzentrationsfähigkeit leidet, weil wir uns über etwas ärgern, oder Angst uns in einer Prüfung besonders leistungsfähig macht. Auch unser Problemlösungsvermögen hängt von unserem Stimmungszustand ab. 1983 konnten Isen und Means nachweisen, dass sich die positiv gestimmten Probanden während eines fiktiven Autokaufs, schneller für ein Auto entschieden, da sie in der Lage waren, die zur Verfügung gestellten Informationen besser zu nutzen [3]. Die „Dunckersche Kerzenaufgabe“ besteht darin, mit Hilfe einer Schachtel Streichhölzer und einer Schachtel Reißnägel, eine Kerze so an der Wand anzubringen, dass kein Wachs auf den Boden tropfen kann. Die Lösung ist, eine der beiden Schachteln als Kerzenhalter zweckzuentfremden. Isen, Daubman und Nowicki gelang es 1987 nachzuweisen, dass die positiv gestimmten Versuchspersonen auf diesen Trick öfter kamen als die neutral Gestimmten, unsere Stimmung also Einfluss auf das kreative Problemlösungsvermögen hat [4]. Aus anderen Studien, die zu ähnlichen Ergebnissen kommen, kann man allgemein ableiten, dass sich eine positive Stimmung vorteilhaft auf unsere Fähigkeit Probleme zu lösen auswirkt. In zwischenmenschlicher Kommunikation spielt v. a. der Ausdruck von Emotionen eine wichtige Rolle. Er hilft unseren Mitmenschen den eigenen emotionalen Zustand mitzuteilen sowie

Handlungsabsichten und Bewertungen zu übermitteln.

2.3 Emotionsmodelle

Ebensowenig, wie man sich bisher auf eine allgemein akzeptierte Definition für die Entstehung von Emotionen einigen konnte, gibt es auch kein einheitliches Emotionsmodell. Die meisten von ihnen basieren auf einem mehrdimensionalen Raum, der von den Basisemotionen aufgespannt wird. Paul Ekman bezeichnet nur solche Emotionen als Basisemotionen, die uns angeboren sind und in jeder Kultur zu finden sind [5]. Die Vorstellung darüber, welche Emotionen man als elementar ansehen kann, gehen jedoch weit auseinander. Tabelle 2.1 zeigt eine Auswahl von Vorschlägen einiger Wissenschaftler [6]. Das 3-dimensionale Emotionsmodell geht

Wissenschaftler	Basisemotionen
Weiner, Graham	Freude, Trauer
James	Angst, Kummer, Liebe, Zorn
Ekman	Freude, Überraschung, Trauer, Angst, Ekel, Ärger
Frijda	Begierde, Freude, Interesse, Überraschung, Verwunderung, Kummer
Arnold	Ärger, Abneigung, Mut, Schwermut, Begierde, Verzweiflung, Furcht, Hass, Hoffnung, Liebe, Trauer

Tabelle 2.1: Basisemotionen verschiedener Wissenschaftler.

auf Wilhelm Wundt (1910) zurück und ordnet die verschiedenen Emotionskategorien entsprechend folgender bipolarer Dimensionen ein [7]:

1. „Lust vs. Unlust“
2. „Erregung vs. Beruhigung“
3. „Spannung vs. Lösung“

Man unterteilt die Emotionen nach dem Maß, in dem sie Lust, Erregung, Spannung und deren Gegenteile enthalten. Somit können alle Emotionen in einem 3-dimensionalen Raum mit orthogonalen Achsen dargestellt werden (Abb. 2.1). Lässt man die dritte Dimension Spannung weg, so erhält man ein 2-dimensionales Emotionsmodell. In Abbildung 2.1 sieht man, wie sich verschiedene Emotionen in diesem Raum abbilden lassen. Man sollte bei dieser Form der Klassifizierung jedoch bedenken, dass die beiden Achsen, auch wenn sie orthogonal zueinander stehen, nicht

zwangsläufig unabhängig sein müssen. Untersuchungen, bei denen Testpersonen verschiedene Emotionen in das System eintragen sollten, ergaben eine ungleiche Verteilung der Emotionen auf die vier Quadranten. Der dritte Quadrant wurde z. B. auffällig selten gewählt [8].

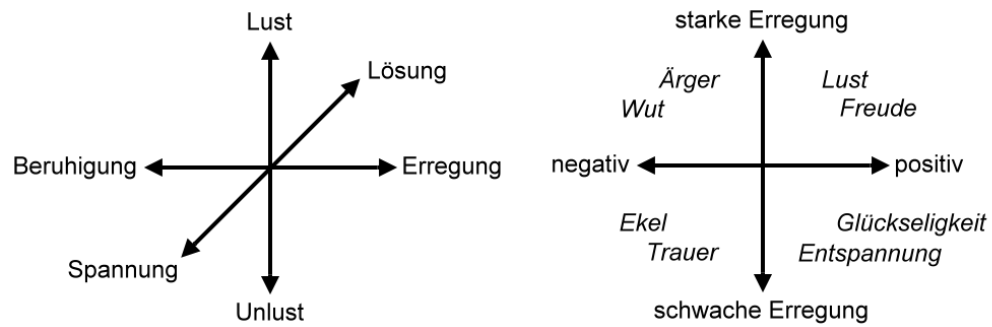


Abbildung 2.1: Die Achsen des 2- und 3-dimensionalen Emotionsmodelles.

Kapitel 3

Affective Computing

In der zwischenmenschlichen Kommunikation spielen Emotionen eine große Rolle. Aus Körperhaltung, Gestik und Mimik anderer Menschen können wir uns ein Bild über deren emotionalen Zustand machen. Nuancen in der menschlichen Sprache teilen uns mit - unabhängig von der Semantik des Gesprochenen - wie sich der Sprecher fühlt. Die Fähigkeit, den physiologischen Zustand unserer Mitmenschen abzuschätzen ist wichtiger Bestandteil natürlicher Kommunikation. Das Bestreben, dem Computer ebenfalls Zugang zu diesen Informationen zu geben und damit die Möglichkeit einer neuen, natürlicheren Interaktion zwischen Mensch und Maschine zu schaffen, die der uns vertrauten zwischenmenschlichen Kommunikation ähnelt, bildet die Grundlage für „Affective Computing“. Im Affective Computing bekommt der Computer die Fähigkeit, Emotionen zu erkennen, eigene Emotionen auszudrücken und dieses Wissen zu einer intelligenten Interaktion mit dem Anwender zu nutzen [22]. Dieses Kapitel beschäftigt sich mit der Frage, welche neuen Möglichkeiten sich damit für die Human-Computer-Interaction (HCI) eröffnen. Außerdem gewährt es einen Einblick inwieweit Affective Computing bereits Gegenstand der Forschung ist, welche Anwendungen schon entwickelt wurden und welche Strategien es gibt, um die Emotionen des Benutzers zu erkennen.

3.1 Neue Möglichkeiten der Interaktion

Die erste kommerzielle Maus wurde am 27. April 1981 auf der National Computer Conference in Chicago vorgestellt und war im Lieferumfang des Xerox Star 8010 enthalten. Zusammen mit der Tastatur ist sie bis heute das wichtigste Eingabegerät, mit dem wir dem Computer unsere Anweisungen mitteilen. Sein Verhalten hängt allein davon ab, welche Befehle er vom Benutzer bekommt, nicht jedoch davon, wie sich dieser fühlt. Dies entspricht nicht der Art von Kommunikation wie wir sie von unseren Mitmenschen gewohnt sind. Wenn wir merken, dass es jemandem

nicht gut geht, bemühen wir uns, alles zu vermeiden, was sich noch zusätzlich negativ auf seine Stimmung auswirken könnte. Wenn wir sehen, dass jemand gestresst ist, werden wir ihn nicht noch mit zusätzlichen Aufgaben belasten. Die Kommunikation zwischen Mensch und Maschine findet jedoch auf einer Ebene statt bei der Gefühle keine Rolle spielen. Hier setzt Affective Computing an. Eine Maschine, die sensibel gegenüber unseren Gefühlen ist, schafft gänzlich neue Möglichkeiten der Interaktion. Sie bietet die Grundlage für eine natürliche, uns vertraute Kommunikation. So lassen sich Hemmschwellen, die viele Menschen im Umgang mit dem Computer haben, abbauen. Ein virtueller Agent etwa, der die Fähigkeit besitzt unsere Gefühle wahrzunehmen und eigene Emotionen durch Sprache, Haltung und Gesten auszudrücken, erscheint nicht mehr bloß als eine kalte, emotionslose Maschine. Der Benutzer kann sich mit ihm identifizieren. Akzeptanz und Vertrauen steigen. Ein solcher Agent ist mehr als ein bloßes „Arbeitsgerät“: er wird zum persönlichen Assistenten und Vertrauten, der auf unsere Bedürfnisse und Wünsche eingeht [9][10].

3.2 Forschung und Anwendung

Erst seit einigen Jahren wird Forschung auf dem Gebiet des Affective Computing betrieben. Vorreiter ist dabei das Massachusetts Institute of Technology (MIT), das eine ganze Reihe von Ideen und Anregungen zu diesem Thema auf seiner Webseite präsentiert [11]. Die Forschung des MIT reicht dabei von der Entwicklung neuer tragbarer Sensoren, über Mechanismen zur Emotionserkennung bis hin zu ersten affektiven Anwendungen, mit denen der Benutzer interagieren kann.

3.2.1 Tragbare Systeme

Die Idee hinter tragbaren Systemen ist den Benutzer auch außerhalb seines Arbeitsplatzes zu jeder Zeit und an jedem Ort zu unterstützen. Tragbare Systeme haben eine eigene CPU und Festplatte. Die eigene Stromzufuhr garantiert ihren mobilen Einsatz. Über verschiedene Ein- und Ausgabeschnittstellen kommunizieren sie mit dem Benutzer. Abbildung 3.1 zeigt Beispiele für Sensoren, die direkt am Körper getragen werden. Die gesammelten Daten können z. B. von einem Palm weiterverarbeitet werden. Um den Benutzer nicht zu behindern, versucht man die nötige Hardware in Kleidung oder Schmuck zu verstecken. Idealerweise wird der Benutzer durch das tragbare System bei alltäglichen Aufgaben unterstützt, in seinem natürlichem Verhalten jedoch nicht gestört. Sog. „Affective Wearables“ sind tragbare Systeme, die darauf spezialisiert sind das emotionale Verhalten des Benutzers aufzuzeichnen und Rückschlüsse auf dessen Zustand zu machen.

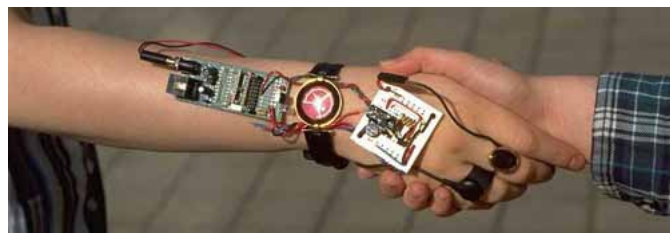


Abbildung 3.1: Wearabels sind mobile Sensoren, die direkt am Körper getragen werden können und z. B. in Alltagsgegenstände wie Kleidung und Schmuck integriert werden. (Quelle: [12]).

3.2.2 Emotionserkennung

Um Emotionen für die Mensch-Maschine-Kommunikation nutzen zu können, muss der Computer in der Lage sein, sich ein Bild von der Stimmung des Benutzers zu machen. Die Individualität mit der jeder Mensch Emotionen erlebt und zeigt, macht es uns oft schwer, den affektiven Zustand unserer Mitmenschen, oder sogar uns selbst, richtig einzuschätzen. Häufig versuchen wir, unsere Stimmung zu verbergen oder eine andere vorzuspielen. Eine Missinterpretation durch den Computer macht die Vorteile eines affektiven Systems zunichte und wird statt einer Hilfe, zu einer Last für den Benutzer. Im folgenden Kapitel werden Methoden vorgestellt werden, die in Zukunft helfen sollen, ein zuverlässiges emotionales Bild des Benutzers zu erstellen.

Gesichtserkennung

Über die Mimik unseres Gesichts offenbaren wir nach außen, wie wir uns im Inneren fühlen. Wir verziehen das Gesicht vor Ekel und lächeln, wenn wir uns freuen. Wir machen ein finsternes Gesicht, weil wir uns über etwas ärgert und runzeln die Stirn, wenn wir nachdenken. Verschiedene Muskeln ermöglichen unserem Gesicht diese Zustände auszudrücken. Durch Verfolgung von festen Referenzpunkten, z. B. entlang der Mundpartie, versucht man aus dem Gesichtsausdruck einer Person auf

ihren emotionalen Zustand zu schließen. (Abb. 3.2).

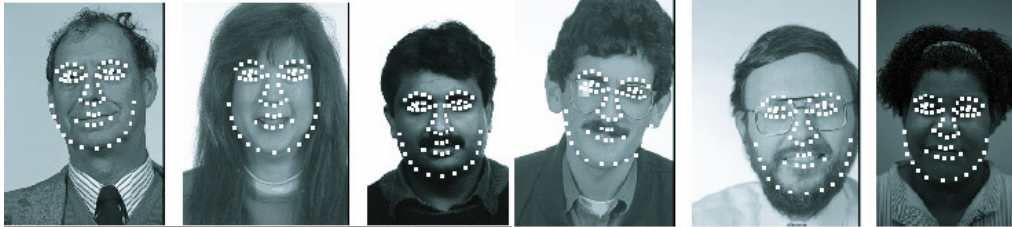


Abbildung 3.2: Gesichtserkennung durch Verfolgung von Referenzpunkten (Quelle: [13]).

Spracherkennung

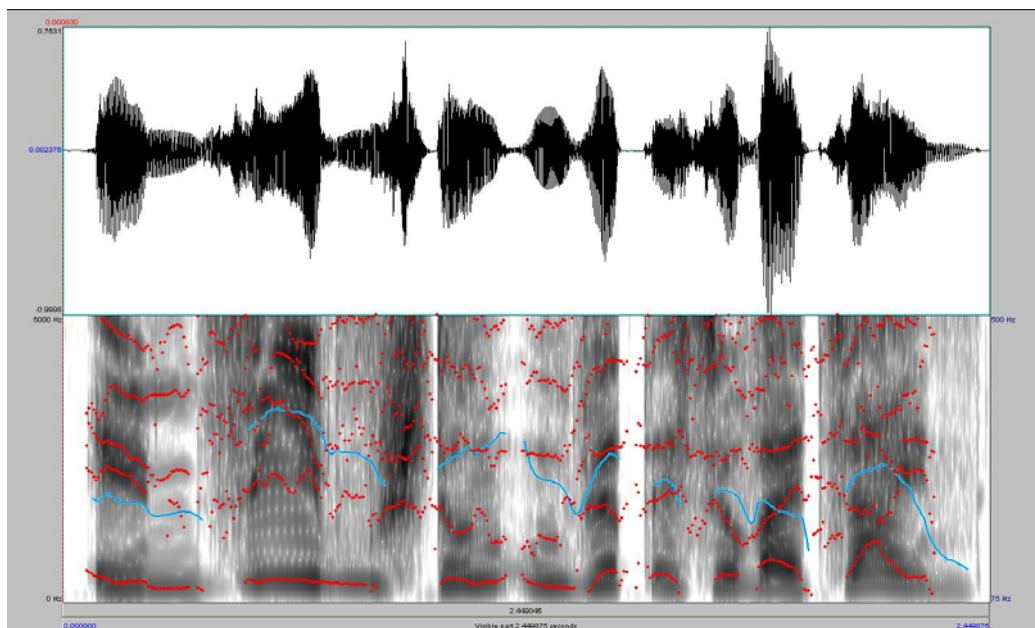


Abbildung 3.3: Feature-Extraktion aus einem Sprachsignal.

Auch im Klang unserer Sprache, unabhängig von der Semantik des Gesagten, spiegelt sich unsere Stimmung wieder. Wir werden leise und eintönig wenn wir traurig sind, schreien wenn wir wütend werden, oder stottern vor Aufregung. Dabei verändern sich in unserer Sprache bestimmte Kennwerte wie Pitch, Harmonics oder Formanten. Mit dem Computer lassen sich diese Kennwerte extrahieren, analysieren und klassifizieren (Abb. 3.3).

Physiologische Daten

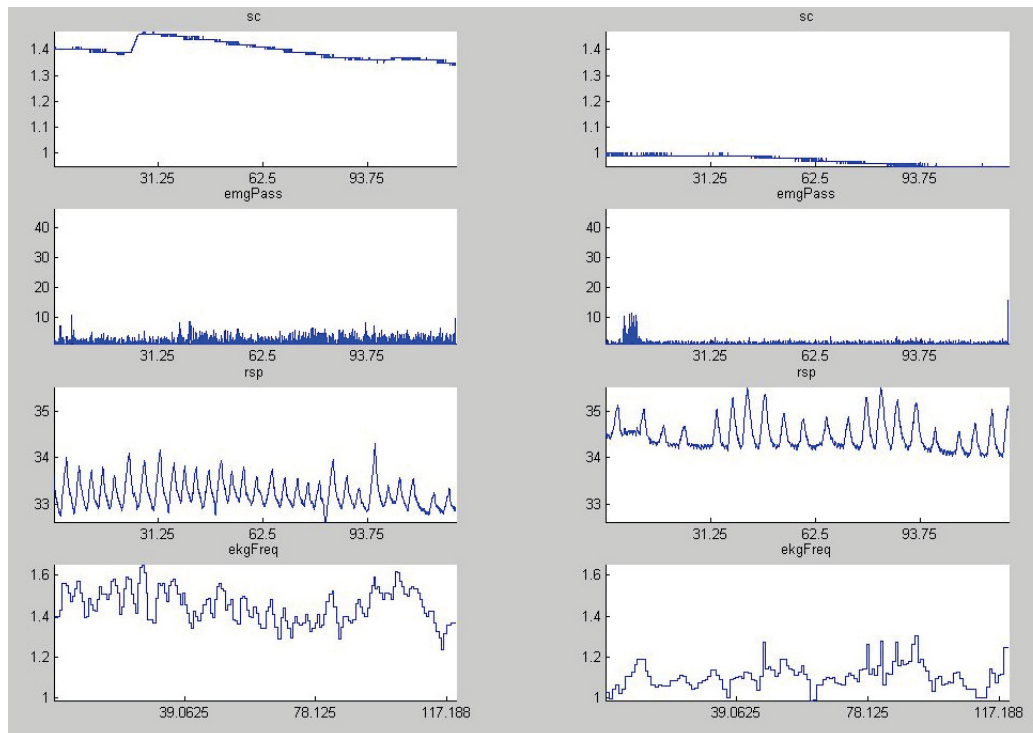


Abbildung 3.4: Hautleitwert, Muskelkontraktion, Atmung und Herzschlag für zwei Aufnahmen mit den Emotionen Freude (l.) und Trauer (r.).

Rückschlüsse auf emotionale Zustände eines Benutzers kann man schließlich auch durch Auswertung seiner physiologischen Daten gewinnen. Elektroden können dazu eingesetzt werden, um z. B. die Schweißproduktion in den Fingerkuppen oder die Kontraktion der Nackenmuskeln zu beobachten, den typischen Boten von Stress. Eine unregelmäßige und schnelle Atmung, die ein Bauchgurt misst, lässt vermuten, dass der Benutzer sich ärgert. Ein schneller Herzschlag deutet daraufhin, dass er erregt ist (Abb. 3.4).

3.2.3 Affektive Anwendungen

Kennt ein System den aktuellen emotionalen Zustand des Benutzers, ergeben sich eine ganze Reihe neuer Anwendungen und Interaktionsmöglichkeiten. Affektive Information kann helfen Routinearbeiten zu erledigen, nächste Arbeitsschritte vorzuschlagen, das Interaktionsverhalten des Geräts anzupassen, oder sogar Einfluss auf den emotionalen Zustand des Benutzers zu nehmen und diesen gegebenenfalls zu verändern.



Abbildung 3.5: Die Abbildung zeigt einen herkömmlichen Desktop, der um affektive Objekte erweitert wurde (Quelle: [15]).

Frustrationsreduktion

Programmabstürze, Fehlverhalten, komplizierte Bedienung oder ein komplexes Interface und Interaktionsverhalten sind häufig Grund für Frustration beim Benutzer. Heutige Systeme haben keine Möglichkeiten, den Benutzern vor solchen Situationen effektiv zu schützen. Ein System dagegen, das sensibel gegenüber Emotionen des Anwenders ist, kann Anzeichen für Frustration frühzeitig erkennen (z. B. wenn der Benutzer gestresst ist oder beginnt sich zu ärgern) und Gegenmaßnahmen ergreifen [14].

Affektives Interface

An der „National University of Singapore“ macht man sich Gedanken, wie ein intelligentes Interface, das Zugang zu physiologischen Informationen über den Benutzer hat, aussehen könnte. In dem Projekt PLASTIC präsentieren Eu Jin Teoh und Daniel Seet ihre Ideen und Entwürfe für eine solche Benutzerschnittstelle [15]. HCI auf Basis von Emotionen könnte einen entscheidenden Ansatzpunkt bieten, um die Produktivität im Umgang mit dem Computer zu steigern. Eine Studie von Dryer und Horowitz (1997) hat gezeigt, dass ähnliche und sich ergänzende Persönlichkeiten besonders gut zusammenarbeiten. Ein Computer, der den emotionalen Zustand seines Benutzers kennt, kann versuchen die Interaktion an dessen Persönlichkeit anzupassen und so ein effizienteres Arbeiten ermöglichen. Dies könnte in Form einer zusätzlichen Schicht zwischen dem Betriebssystem und dem Be-

nutzer realisiert werden. Abbildung 3.5 zeigt einen herkömmlichen Desktop, der um verschiedene affektive Elemente erweitert wurde. Wie Laurel in „Computers as Theatre” beschreibt, ist das Interface die entscheidende Schnittstelle für die Erfahrungen, die wir mit dem Computer machen [16]. Je natürlicher und intuitiver wir diese Schnittstelle nutzen können, desto einfacher und schneller finden wir uns damit zurecht und desto besser sind wir in der Lage sie effektiv zu nutzen. Wer lange zusammenarbeitet lernt auf derselben Ebene zu denken und zu handeln. Nur noch ein Minimum an Kommunikation ist dann zur Verständigung nötig. Ein Computer, der auf ähnliche Weise empfindsam wird für Bedürfnisse seines Benutzers, könnte dieses Prinzip menschlicher Kommunikation erlernen.

Affektive Applikationen

Erste Prototypen von Produkten, die uns vielleicht bald in unserem alltäglichen Leben unterstützen, wurden am MIT entwickelt. Hinter dem „Affective DJ” verbirgt sich ein tragbarer PalmPilot, der aus einer Datenbank Musikstücke passend zu der Stimmung des Benutzers auswählt [17]. Während der Lernphase sammelt er dafür zunächst Informationen über die Präferenzen des Benutzers und vergleicht physiologische Daten, wie Atemfrequenz und Hautleitwert, mit den gespielten Songs. So lernt er die Stimmungen des Benutzers mit bestimmten Musikkategorien zu verknüpfen. Hat er genügend Wissen gesammelt, ist er in der Lage selbstständig die Musikauswahl zu treffen. Abbildung 3.6 zeigt das Interface des PalmPilots und wie die Sensoren am Körper des Benutzers angebracht werden. Affective Computing

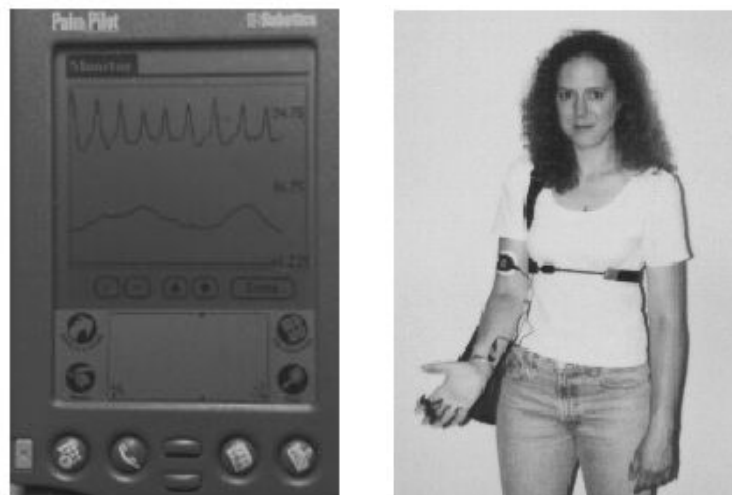


Abbildung 3.6: Der „Affective DJ” lernt Musikstücke entsprechend der emotionalen Stimmung seines Benutzer auszuwählen (Quelle: [17]).

kann aber nicht nur in der HCI eingesetzt werden, sondern auch als Erweiterung der zwischenmenschlichen Kommunikation. Galvactivator heißt ein, ebenfalls am MIT entwickelter Handschuh, der die aktuelle Stimmung seines Trägers anzeigt [18]. Ein weißes Licht bedeutet, dass der Benutzer entspannt ist. Es ändert jedoch seine Farbe und leuchtet rot, wenn er gestresst ist. Durch einen Blick auf den Handschuh können unsere Mitmenschen sofort erkennen, wie gestresst oder entspannt wir gerade sind und sich entsprechend darauf einstellen.



Abbildung 3.7: Der „Galvactivator“ zeigt mit Hilfe einer Leuchtdiode den Stresszustand seines Benutzers an [18].

Kapitel 4

Emotionserkennung mit physiologischen Daten

Die Psychologen Carl Gustav Jung und Max Wertheimer gelten als die Erfinder des sog. Polygraph, auch bekannt als Lügendetektor. Polygraph bedeutet eigentlich Mehrkanalschreiber und misst wie sich Atmung, Blutdruck, Hautwiderstand und Puls eines Probanden bei der Befragung durch einen Psychologen verändern. Sein Einsatz als Beweismittel in gerichtlichen Verfahren ist umstritten. In Deutschland wurde er deshalb 1954 durch einen Spruch des Bundesgerichtshofs als Beweismittel ausgeschlossen. In den USA dagegen ist sein Einsatz bis heute erlaubt. Unabhängig von dieser Kontroverse deutet die Entwicklung des Polygraphen darauf hin, dass sich der affektive Zustand eines Menschen in physiologischen Reaktionen widerspiegelt. Mit dem Affective Computing wurde ein Forschungsgebiet geschaffen, in dem die Auswertung physiologischer Daten eine neue und interessante Bedeutung erhält. Im weiteren Verlauf beschäftigt sich diese Arbeit deshalb mit der Frage, wie physiologische Daten in Zukunft in der Emotionserkennung eingesetzt werden können. Dabei wird in diesem Kapitel zunächst ein Überblick darüber gegeben, welche relevanten physiologischen Daten es gibt und wie sie gemessen werden.

4.1 Motivation

Es folgt ein Überblick über Arbeiten, in denen bereits die Emotionserkennung aus physiologischen Daten untersucht wurde. Anschließend werden Schwierigkeiten und Grenzen, aber auch die Vorteile gegenüber herkömmlichen Verfahren beschrieben.

4.1.1 Verwandte Arbeiten

Bis heute herrscht Streit unter den Psychologen, ob Emotionen mit individuellen physiologischen Mustern verknüpfbar sind. Noch gibt es keine Beweise für die Existenz emotionsspezifischer Muster. Erste Untersuchungen von physiologischen Daten mit Hilfe digitaler Signalverarbeitungstechniken in Kombination mit Mustererkennungsalgorithmen liefern jedoch Erfolg versprechende Ergebnisse. Hervorzuheben sind dabei die Arbeiten von Jennifer Healey [19] und Elias Vyzas [20] am MIT, die gleichzeitig auch die Grundlage für diese Arbeit bilden. Für ihre Untersuchungen zeichneten sie physiologische Daten in insgesamt acht verschiedenen emotionalen Zuständen auf. Mit Hilfe statistischer Analyseverfahren konnten die aufgezeichneten Emotionen anschließend mit Wahrscheinlichkeiten deutlich höher als Zufall richtig zugeordnet werden. Lisetti und Nasoz entwickelten an der Universität von Florida ein affektives Interface (Multimodal Affective User Interface, MAUI), das in Zukunft an spezielle Applikationen angepasst werden kann [30]. Neben Kanälen für Gesicht und Sprache erhält das Interface als zusätzlichen Input auch die physiologischen Daten des Benutzers. Für die Erforschung des dritten Kanals erstellten sie ebenfalls einen Datenkorpus mit Biodaten und konnten mit statistischen Klassifikationsverfahren am Ende sechs Emotionen mit einer Wahrscheinlichkeit von über 80% richtig zuordnen. Die Autoren von [27] beschreiben den Einsatz eines neuronalen Netzes zur Emotionserkennung aus physiologischen Daten und erzielten damit ähnlich gute Ergebnisse.

4.1.2 Schwierigkeiten und Grenzen

Nun gibt es durchaus einige Gründe, die erklären, warum es so schwer fällt, die Reaktionen unseres Körpers eindeutig einem emotionalen Zustand zuzuordnen [22]. Dieselbe Emotion können wir je nach Intensität auf sehr unterschiedliche Weise empfinden. Eine emotionale Reaktion auf ein uns bisher unbekanntes Erlebnis erfahren wir viel eindringlicher. Je öfter wir jedoch mit derselben Situation konfrontiert werden, desto schwächer reagiert unser Körper darauf (z. B. auf Szenen in einem Film, den man schon mehrmals gesehen hat). Auch die Umgebung, in der wir uns gerade befinden, hat Einfluss darauf (z. B. macht es einen Unterschied, ob man bei Tag oder Nacht allein durch einen Wald geht). Eine körperliche Reaktion muss auch nicht immer zwingend mit dem Erleben einer Emotion verknüpft sein (z. B. bei sportlicher Aktivität). Weiter gibt es eine ganze Anzahl von äußeren Faktoren: z. B. Alter und Fitness, aber auch das kulturelle Umfeld oder die Erfahrungen und das Wissen der Person. Schließlich reagiert jeder Mensch auf dieselben Emotionen mehr oder weniger individuell, was die Suche nach gemeinsamen physiologischen Mustern sehr schwer macht. Genau nach solchen allgemeingültigen Regeln haben Emotionsforscher bisher aber meistens gesucht. Ein adaptives System, das sich auf

seinen jeweiligen Benutzer anpasst und mit der Zeit ein persönliches Bild von ihm entwirft, hat wesentlich bessere Aussichten auf Erfolg. Zugang zu Kontextinformation und die Kombination mit Gesichts- und Spracherkennern stellen eine weitere Möglichkeit zur Verbesserung dar. Der Erforschung von hybriden Systemen wird in Zukunft eine sehr wichtige Bedeutung zukommen.

4.1.3 Vorteile gegenüber herkömmlichen Verfahren

Physiologische Daten zur Emotionserkennung zu verwenden, ist im Gegensatz zur Gesichts- und Spracherkennung erst seit wenigen Jahren ein Thema in der Forschung. Dabei gibt es einige Vorteile gegenüber den herkömmlichen Methoden [23]. Unser Körper sendet die gemessenen Signale ohne Unterbrechung, so dass man einen kontinuierlichen Signalstrom erhält. Da der Benutzer zudem die Sensoren direkt am Körper trägt, kann es nicht passieren, dass er aus ihrer Reichweite gerät. Mobile Sensoren können dann die Aufzeichnung in jeder beliebigen Umgebung fortführen: z. B. in der Arbeit, beim Einkaufen oder während der Autofahrt. Durch die Entwicklung von Sensoren, die in Kleidung oder Schmuck integriert sind (vgl. Kapitel 3.2.1), wird der Benutzer in seinem natürlichen Verhalten nicht gestört und kann sich in gewohnter Weise verhalten. Während wir Mimik und Sprache relativ gut kontrollieren und damit bewußt oder unbewußt verstellen können, werden die physiologischen Reaktionen unseres Körpers durch unser zentrales Nervensystem gesteuert und sind nur teilweise oder gar nicht der Kontrolle unseres Bewusstseins unterlegen. Es besteht so eine direkte und unverfälschte Verbindung zu unseren Emotionen. Auch in der zwischenmenschlichen Kommunikation spielen physiologische Daten eine oft unterschätzte Rolle. Wenn wir jemanden schnell und unregelmäßig schnaufen hören, vermuten wir, dass er gestresst ist. Ein feuchter Händedruck verrät uns Erregung. Mimik schließlich ist nichts anderes als das Ergebnis von Muskelkontraktionen in unserem Gesicht. Manchen Menschen erscheint die Aufzeichnung physiologischer Daten zunächst als ein zu starker Eingriff in die Privatsphäre. Im Gegensatz zu Videoaufzeichnungen werden Identität, Auftreten und Verhalten aber sogar besser geschützt.

4.2 Sensoren und Signale

Biosignale sind alle physikalisch messbaren und kontinuierlich oder nahezu kontinuierlich registrierbaren Körperfunktionen. Hierbei unterscheidet man direkte bioelektrische Signale (z.B. Herzschlag, Hirnaktivität), indirekte bioelektrische Signale (z.B. Hautleitfähigkeit) und nicht elektrische Signale (z.B. Blutdruck, Atemfrequenz). In der Medizin setzt man seit langem Sensoren ein, um diese Signale zu messen. Aus den Aufzeichnungen kann man Rückschlüsse auf die körperliche

Verfassung des Patienten ziehen und sogar lebensbedrohliche Krankheiten (z. B. Herzfehler) erkennen. Für viele Sensoren werden mittlerweile digitale Varianten angeboten. Für diese Arbeit wurde der ProComp+-Encoder von Thought Technology Ltd. verwendet [24]. Vier Sensoren zur Messung von Hautleitwert, Muskelkontraktion, Atmung und Herzschlag standen zur Verfügung. Es folgen Erklärungen zu den einzelnen Sensoren und Signalen. Die Ausführungen nehmen im Wesentlichen Bezug auf das „Lehrbuch Psychologie“ von Rainer Schandry [25].

4.2.1 Hautleitwert (SC)

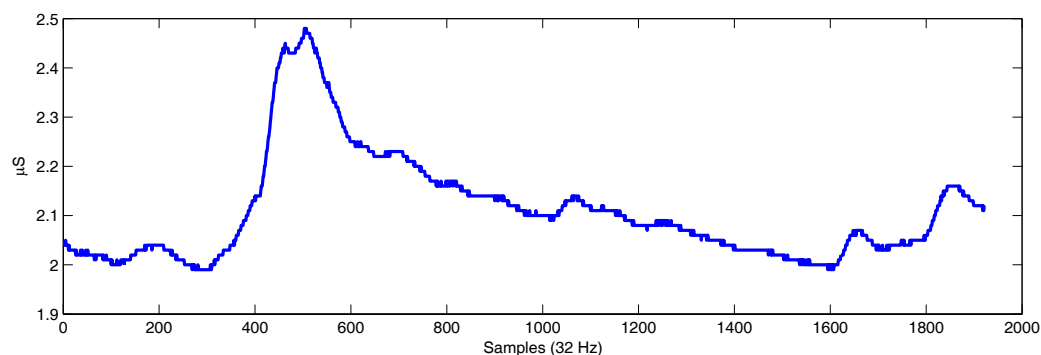


Abbildung 4.1: Der Verlauf eines typischen SC-Signals.

Anbringung und Messung

Der Hautleitwert (SC) misst die elektrische Leitfähigkeit der Haut (Abb. 4.1). Diese wird von der Aktivität der Schweißdrüsen beeinflusst. Die dichteste Verteilung der Schweißdrüsen findet sich an den Hand- und Fußinnenseiten. Die Formen simultan aufgezeichneter Signale an Hand und Fuss ähneln sich stark, wobei an der Hand mit etwas deutlicheren Ergebnissen zu rechnen ist [19]. Bei den Aufzeichnungen für den Datenkorpus wurde je eine Elektrode am linken Zeige- und Mittelfinger angebracht (Abb. 4.2). Zwischen diesen beiden Elektroden wird an der Hautoberfläche eine geringe Spannung angelegt und der Hautleitwert gemessen. Bei Schweißproduktion steigt die Fähigkeit der Haut als Stromleiter zu fungieren und dementsprechend auch der Hautleitwert an. Schweißdrüsen spielen bei der Erfassung von Veränderungen bioelektrischer Eigenschaften der Haut eine sehr wichtige Rolle. An den Elektroden wird eine konstant bleibende Spannung (0,5 Volt, nach Edelberg (1967)) angelegt. Der aus dem gemessenen Strom gewonnene Hautleitwert trägt die Maßeinheit Mikrosiemens (μS), wobei Siemens dem Reziproken von Ohm (Ω^{-1}) entspricht. Da der Leitwert eines bestimmten Materials oder Gewebes proportional



Abbildung 4.2: Anbringung der Elektroden zur Aufzeichnung des SC-Signals an den Fingerkuppen der Versuchsperson.

zum Querschnitt zunimmt, sind Zahlenwerte der Leitfähigkeit stets auf die Fläche (hier auf den Elektrodenquerschnitt in cm^2) zu beziehen.

Deutung des Signals

Besonders sensibel reagiert der Hautleitwert auf Stress und Erregung. Reaktionen des Hautleitwerts können durch drei Kennwerte charakterisiert werden. Die Amplitude des Ausschlags relativ zur Grundlinie, die Zeit bis zum Erreichen des maximalen Ausschlags und schließlich die Dauer bis zum erneuten Absinken auf das Grundniveau. Da es schwierig ist zu antizipieren wann das Signal wieder auf das Grundniveau abgesunken ist, kann auch die Dauer gemessen werden, die verstreicht bis der Ausschlag wieder um die Hälfte zurückgegangen ist. Hat man zusätzlich noch Wissen über den Grund des Ausschlags, so kann man als vierten Kennwert noch die Latenz zwischen Stimulus und Reaktion heranziehen. Diese ist insofern interessant, als der Hautleitwert stets leicht versetzt erst nach etwa 1-2 Sekunden einsetzt. Eine zusätzliche Schwierigkeit bei der Auswertung ergibt sich, wenn während des Abfalls des Signals auf das Grundniveau ein zweiter Ausschlag des Hautleitwerts erfolgt. Insgesamt ist der Hautleitwert ein zuverlässiges Maß für Aktivitäten des menschlichen Nervensystems. Da die meisten Emotionen eine erhöhte Aktivität unseres Nervensystems zur Folge haben, ist ein Ausschlag des Hautleitwerts ein sicheres Zeichen dafür, dass eine Emotion vorliegt. Die Amplitude und Dauer der Reaktion sind Maßstab für die Intensität und Dauer der Emotion.

Messfehler und Artefakte

Durch verschiedene innere oder äußere Einflüsse können Signale, die von keinem psychophysiologischen Interesse sind, an die Elektroden weitergeleitet werden. Beispielsweise lässt eine erhöhte Körpertemperatur auch den Hautleitwert steigen. Da die Elektroden direkt an der Haut anliegen erwärmt sich die Haut unter ihnen, was den gleichen Effekt hat. Damit die Ergebnisse nicht verfälscht werden, wird später bei der Auswertung ein Algorithmus verwendet, der den durch diesen Umstand bedingten unvermeidlichen Anstieg des Hautleitwertniveaus ausgleicht. Ein weiteres zu beobachtendes Phänomen ist der Anstieg des Hautleitwerts, wenn die angeschlossene Person tiefe Atemzüge macht oder lange die Luft anhält, z. B. wenn sie spricht. Bewegungsartefakte treten auf, sobald die Finger oder die Hand mit den Elektroden stark bewegt werden.

4.2.2 Muskelkontraktion (EMG)

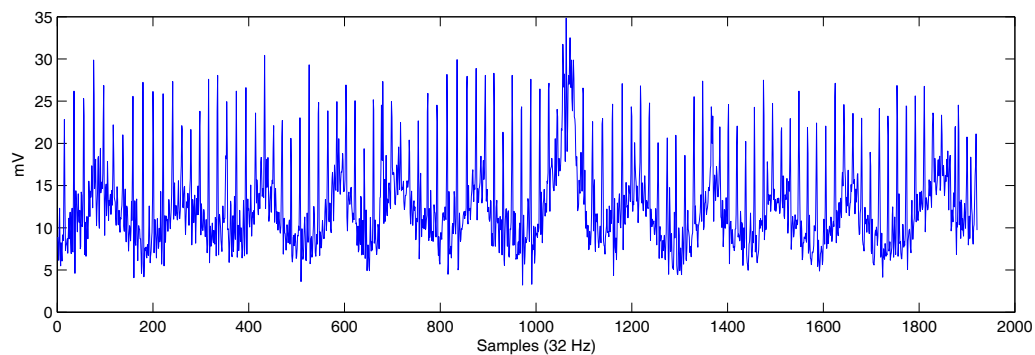


Abbildung 4.3: Der Verlauf eines typischen EMG-Signals.

Anbringung und Messung

Das Elektromyogramm (EMG) dient zur Erfassung der elektrischen Muskelaktivität (Abb. 4.3). Das Verfahren, das zur Darstellung dieser Körperfunktion verwendet wird, nennt man (Oberflächen-) Elektromyographie. Die Amplitude dient als Intensitätsanzeige. Zur Messung werden drei Elektroden verwendet, die prinzipiell eingesetzt werden können um einen beliebigen Muskel zu messen. In den Aufzeichnungen für den Korpus wurden die Elektroden im Nacken der Versuchsperson angebracht (Abb. 4.4). Auch die Stirn- oder Backengegend sind typische Positionen für EMG-Messungen. Je nach Anbringung können die Ergebnisse bei simultanen Aufzeichnungen jedoch stark voneinander abweichen. Das EMG-Signal resultiert

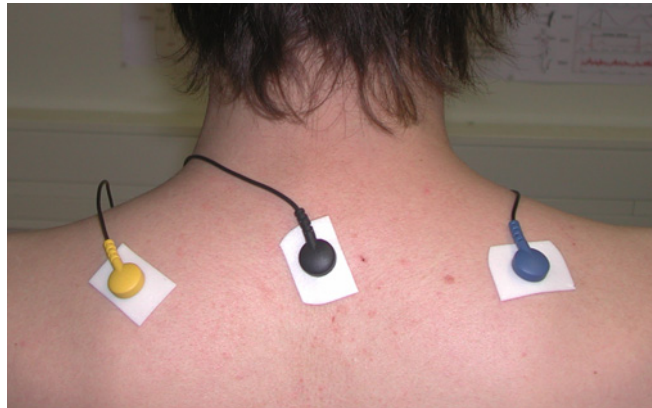


Abbildung 4.4: Anbringung der Elektroden zur Aufzeichnung des EMG-Signals im Nacken der Versuchsperson.

aus additiven Überlagerungen vieler einzelner Aktionspotentiale und verschiedener Störsignale. Ihre Amplituden betragen einige Mikrovolt bis etwa zehn Millivolt. Das Frequenzspektrum reicht von ca. 5 bis 400 Hz, bei „schnellen“ Muskeln bzw. solchen mit hoher Feinkoordination bis etwa 700, maximal 800 Hz. Üblicher Abstand der Messelektroden ist bei bipolarer Messung 25 bis 30 mm. Für spezielle Aufgabenstellungen kann jedoch auch ein geringerer oder größerer Abstand erforderlich sein. Sowohl die aktive Fläche der Elektroden, als auch der Elektrodenabstand beeinflussen das gemessene Signal, insbesondere das Frequenzspektrum. Bei vergleichenden Untersuchungen sollte man deshalb gleiche Elektroden verwenden und auf gleiche Anbringung achten.

Deutung des Signals

Das EMG misst Muskelkontraktionen, d. h. der Ausschlag ist stärker, wenn der Muskel angespannt wird und fällt bei Entspannung. Stress oder starke Konzentration führen zu Verkrampfungen bestimmter Muskelregionen, z. B. am Rücken oder im Nacken. Werden die Elektroden dagegen an der Stirn angebracht, beobachtet man Ausschläge, wenn die Stirn gerunzelt oder gestreckt wird. Wie das Signal gedeutet werden muss hängt also stark von der Platzierung der Sensoren ab. Abgesehen davon kann das EMG auch zur Erkennung von Mimik oder Gestik eingesetzt werden [26].

Messfehler und Artefakte

Großen Einfluss auf die Signalqualität hat die Hautvorbereitung. Auf der Haut befindet sich eine Schicht aus abgestorbenen Hautzellen. Diese Schicht besitzt einen

hohen elektrischen Widerstand, der die Messungen erheblich negativ beeinflussen kann. Zusätzlich können Hautcremes zur Erhöhung des Übergangswiderstandes beitragen und verschlechtern die Haftung der selbstklebenden Elektroden auf der Haut. Die Hautstellen, an denen Elektroden angebracht werden, sollten deshalb vorher mit einem Wattepad und Desinfektionsmittel gereinigt werden, um möglichst Artefakte zu vermeiden und ein deutlicheres Signal zu erhalten. Bei Anbringung der Elektroden im Nacken wird das Signal durch das leichte Heben und Senken der Schultern während des Atemvorgangs gestört. Zusätzlich finden sich regelmäßige scharfe Ausschläge, die der Herzfrequenz entsprechen, wieder. Beide Artefaktquellen müssen bei der Auswertung durch geeignete Algorithmen aus dem Signal entfernt werden. Zusätzlich wird das Signal durch Bewegung des Kopfes oder des Oberkörpers, insbesondere der Schultern, gestört.

4.2.3 Atmung (RSP)

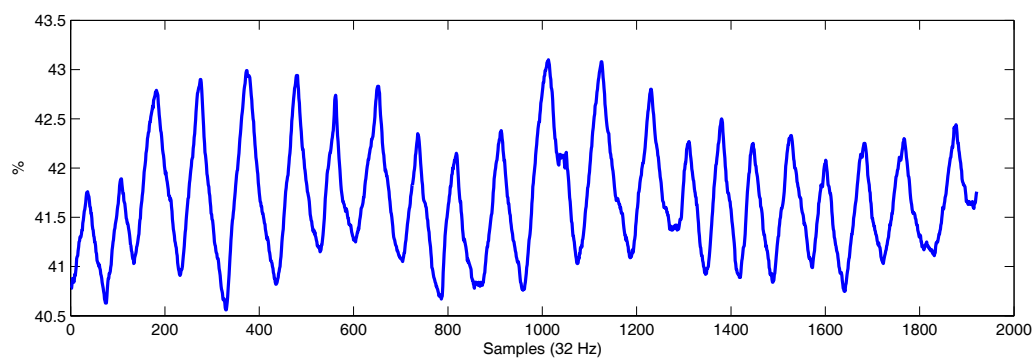


Abbildung 4.5: Der Verlauf eines typischen RSP-Signals.

Anbringung und Messung

Ein Atemzug umfasst je einmal Ein- und Ausatmen. Der Normalwert der Atemfrequenz ist sowohl alters- als auch belastungsabhängig. Unter Ruhebedingungen holt ein Erwachsener etwa 15 mal pro Minute Luft (Abb. 4.5). Zur Erfassung der respiratorischen Aktivität wird ein dehnungssensitiver Gurt um den Oberkörper der Versuchsperson gespannt. Atmet eine Person über die Brust (Brustatmer), so muss der Gurt im oberen Bereich der Brust angebracht werden. Wird über den Bauch geatmet (Bauchatmer), so muss man den Gurt am Bauch befestigen (Abb. 4.6). Atmet die Person ein, erhöht sich die Spannung auf den Atemgurt. Seine relative Ausdehnung wird von dem am Gurt angebrachten Sensor registriert und aufgezeichnet. Atmet die Person wieder aus, entspannt sich der Atemgurt und der

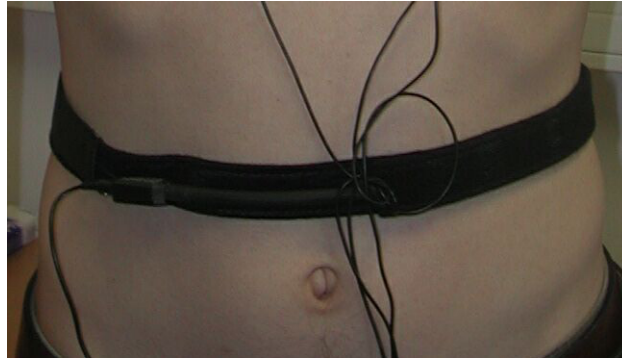


Abbildung 4.6: Anbringung des Bauchgurts zur Aufzeichnung des RSP-Signals.

Sensor kehrt zu seiner neutralen Position zurück. Die Atemkurve wird gemessen und das gespeicherte digitalisierte Rohsignal kann auf dem Bildschirm dargestellt werden.

Deutung des Signals

Ist eine Person entspannt, so ist ihre Atmung gleichmäßig und ruhig. Sie verschnellert sich bei emotionaler Erregung oder physischer Anstrengung. Plötzliche Reize können aber auch ein vorübergehendes Aussetzen der Atmung zur Folge haben (z. B. in Schrecksituationen). Erregung führt zu Unregelmäßigkeiten im Signal. Dies ist auch der Fall wenn die Versuchsperson spricht oder lacht.

Messfehler und Artefakte

Das RSP-Signal ist sehr stark durch Artefakte geprägt. Der Gurt um den Bauch des Probanden kann durch das Heben und Senken der Brust während des Ein- und Ausatmevorgangs leicht verrutschen. Wenn wir sprechen oder lachen setzt unsere Atmung kurzfristig aus. Wir gleichen dies meist aus, indem wir kurz davor und danach besonders tiefe Atemzüge machen. Die Pausen und Phasen starker Inhalation stören die Messung der Atemfrequenz. Zur sinnvollen Auswertung des Atemsignals ist der Einsatz spezieller Frequenzfilter deshalb unumgänglich.

4.2.4 Herzschlag (EKG)

Anbringung und Messung

Das Elektrokardiogramm (EKG) zeichnet Herzströme auf (Abb. 4.7). In Abhängigkeit der Schreibgeschwindigkeit des Aufzeichnungsgerätes kann man die Herzfrequenz, also die Anzahl der Herzschläge pro Minute errechnen. Das Ruhepotential

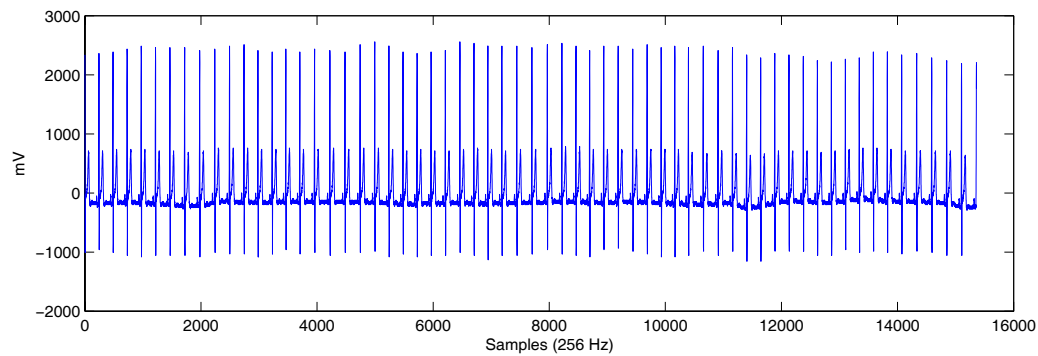


Abbildung 4.7: Der Verlauf eines typischen EKG-Signals.

eines gesunden Menschen - die Anzahl der Herzschläge während einer Entspannungsphase - liegt bei etwa 60-70 Schlägen pro Minute, unter starker Anstrengung kann der Puls auf einen Maximalwert von bis zu 200 Schlägen pro Minute ansteigen. Dies kann jedoch variieren, da jeder Organismus unterschiedliche Grundvoraussetzungen (Geschlecht, Alter, Fitness, usw.) mitbringt. Das EKG-Signal ist periodisch und weist eine Reihe von Merkmalen auf. Am deutlichsten tritt die R-Zacke hervor. Aus der Differenz zweier aufeinanderfolgender R-Zacken lässt sich die Herzfrequenz bestimmen. Eine der vielen Möglichkeiten, die Elektroden anzubringen, zeigt Abbildung 4.8.

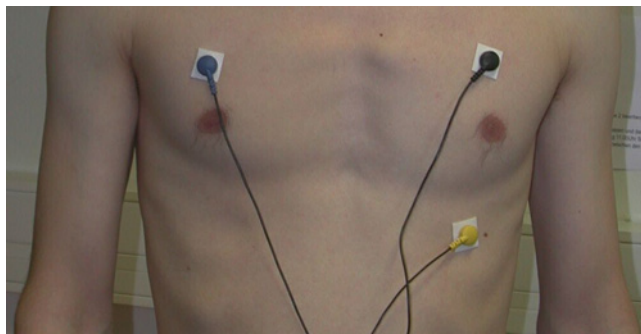


Abbildung 4.8: Anbringung der Elektroden zur Aufzeichnung des EKG-Signals am Körper der Versuchsperson.

Deutung des Signals

In der Literatur wird beschrieben, dass bei physischer Anstrengung (auch starker Konzentration), aber auch Stress und Furcht die Herzfrequenz zunimmt, bei Trauer und Ärger dagegen nimmt sie ab [25].

Messfehler und Artefakte

Bewegungen zählen hierbei zu den häufigsten Artefaktquellen. Außerdem können im EKG-Signal Hautpotential- und Muskelpotentialartefakte überlagert sein. Hautpotentialänderungen machen sich in einer langsamen, wellenförmigen Schwankung der Grundlinie bemerkbar. Muskelpotentiale von Muskelgruppen, die unter den Elektroden liegen, können ebenfalls die Messung stören.

Kapitel 5

Aufbau des Datenkorpus

Um ein System, das Emotionen aus physiologischen Daten erkennen soll, trainieren zu können, benötigt man zunächst einen Korpus, der ausreichend Beispiele für die Emotionen enthält. Im Gegensatz zur Sprach- und Gesichtserkennung, gibt es kaum verfügbare Trainingssets mit biologischen Daten. Für diese Arbeit konnte lediglich auf den am MIT aufgezeichneten und im Internet frei verfügbaren Korpus des „Eight Emotion Experiment“ zurückgegriffen werden [11]. Eine Aufgabe dieser Arbeit bestand deshalb auch aus der Aufzeichnung eines eigenen Korpus. Dafür wurden zunächst im Selbstversuch über mehrere Wochen in einem von Dr. Jonghwa Kim entworfenem Experiment physiologische Daten aufgezeichnet. Der Korpus mit dem Namen AuDB (Augsburger Database of Biosignal) soll in Zukunft um weitere Daten erweitert werden und Grundlage für die eigene Forschung bilden, aber auch anderen Universitäten für ihre Arbeit zur Verfügung gestellt werden. Dieses Kapitel beginnt zunächst mit einigen allgemeinen Überlegungen, die bei der Erstellung der Datenbank berücksichtigt wurden. Anschließend folgt eine Beschreibung des Experiments und der Aufzeichnungen.

5.1 Vorüberlegungen zum Korpus

Die Erstellung eines Korpus ist mit großem Aufwand verbunden. Will man genügend Material sammeln, können sich die Aufnahmen über Wochen oder Monate hinziehen. Es sollte sich deshalb stets um dieselben Versuchspersonen handeln und möglichst ähnliche Aufnahmebedingungen herrschen. Da der Erfolg der späteren Auswertung in starkem Maße von der Qualität des Korpus abhängt, sollte genügend Zeit in die Planung des Experiments verwendet werden. Im Folgenden wird auf verschiedene Punkte eingegangen, die bei der Erstellung des Experiments berücksichtigt wurden.

5.1.1 Auswahl

Die erste Entscheidung betrifft die Auswahl der Emotionen, die das System erlernen soll. Der Erfolg des Systems hängt insbesondere davon ab, wie deutlich sich die Emotionen unterscheiden. Wenn man an das 2-dimensionale Modell aus Kapitel 2.3 zurückdenkt, in dem die Emotionen entsprechend ihrem Anteil an Erregung und Lust eingeteilt wurden, so lässt sich vermuten, dass sich vor allem solche Emotionen gut voneinander unterscheiden lassen, die auch räumlich getrennt sind - insbesondere solche in verschiedenen Quadranten. Die Emotionen Freude, Glückseligkeit, Trauer und Ärger wurden deshalb ausgesucht, um jeden der vier Quadranten abzudecken (Abb. 5.1). Freude, Trauer und Ärger tauchen unter ähnlicher Bezeichnung auch in den Basisemotionen vieler Emotionsforscher wie James, Ekman oder Clynes auf. Glückseligkeit wurde als eine positive Emotion mit schwacher Erregung gewählt.

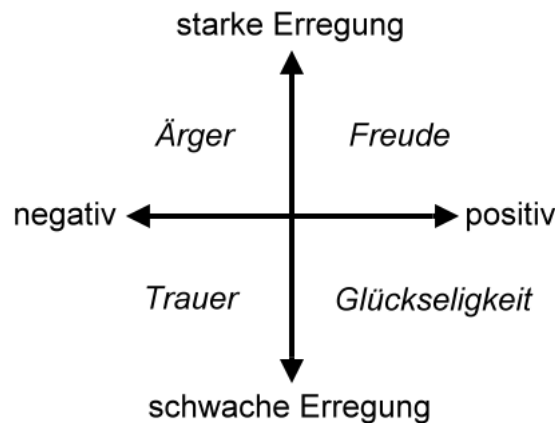


Abbildung 5.1: Das Emotionsmodell für die Aufzeichnung des eigenen Korpus.

5.1.2 Aufnahme

Hat man sich für Anzahl und Art der zu untersuchenden Emotionen sowie das zugrundeliegende Emotionsmodell entschieden, stellt sich die Frage, wie man möglichst echte Aufnahmen erhält. Hierbei unterscheidet man drei Kategorien:

- spontane Emotionen
- gespielte Emotionen
- kontrolliert ausgelöste Emotionen

Spontane Emotionen

Innerhalb dieser drei Kategorien sind die spontanen Emotionen natürlich am authentischsten und eignen sich somit am besten für eine Klassifizierung. Spontane Emotionen setzen jedoch insbesondere voraus, dass die Versuchsperson nicht weiß, dass sie aufgezeichnet wird. Abgesehen von rechtlichen Fragen ist dies beim momentanen Stand der Technik für physiologische Daten kaum möglich, da die Sensoren direkt am Körper getragen werden und per Kabel mit dem Rekorder verbunden sind. Außerdem hat man hier natürlich auch keine Kontrolle darüber, welche Emotion wann und wie lange auftritt. Ein guter Korpus sollte jedoch etwa gleichviele und gleichlange Beispiele für jede Emotion enthalten. Außerdem müssen die Daten anschließend noch gelabelt werden, d.h. das Signal muss in Bereiche zerlegt werden, denen jeweils eine bestimmte Emotion zugeordnet werden kann. Dieser Vorgang wird als Annotation bezeichnet und erfordert in der Regel einen großen Aufwand.

Gespielte Emotionen

Bei gezielten Emotionen versucht sich der Proband künstlich in die Emotionen zu versetzen. Die Reihenfolge und Dauer sind hier exakt vorgegeben, so dass sich der Aufbau des Korpus genau kontrollieren lässt. Allerdings hängt jetzt alles davon ab, wie gut es der aufgezeichneten Person gelingt sich in die Emotionen zu versetzen. Oft fällt Menschen ein natürliches Verhalten unter Aufnahmebedingungen schwer. Für Sprach- und Gesichtserkennung greift man deshalb häufig auf Schauspieler zurück, da sie aufgrund ihres Berufes gewohnt sind Emotionen vorzuspielen. Bei physiologischen Daten genügt es jedoch nicht eine Emotion nur äußerlich durch Ändern der Mimik oder Verstellen der Stimme vorzutäuschen. Die Reaktionen unseres Organismus, wie z. B. der Hautleitwert, werden nämlich durch unser zentrales Nervensystem angeregt, ohne dass wir Einfluss darauf haben.

Kontrolliert ausgelöste Emotionen

Die dritte Methode stellt eine Kombination aus den ersten beiden dar: möglichst spontane Emotionen unter kontrollierten Bedingungen. Für mehr Authentizität versucht man die Versuchsperson in eine Situation zu versetzen, in der es ihr leichter fällt, die gewünschte Emotion zu erleben. Ob der angestrebte Effekt eintritt, kann natürlich nicht garantiert werden. Verschiedene Personen können auf denselben Reiz sehr verschieden reagieren. Zusätzlich stößt man hierbei auf ethische Probleme, wenn man bei den Probanden mit Absicht negative Emotionen erzeugen möchte. Da aufgrund technischer Einschränkungen die Aufzeichnung spontaner Emotionen noch nicht möglich ist und gespielte Emotionen nicht authentisch genug sind, stellen kontrolliert ausgelöste Emotionen dennoch die beste Alternative für

physiologische Daten dar. Auch für diesen Korpus fiel die Wahl deshalb auf diese Methode.

5.1.3 Erregung

Es gibt viele Möglichkeiten, um Emotionen zu stimulieren. Die Autoren von [27] benutzten für ihr Experiment Fotos aus dem IAPS¹, das 800 Fotografien, sortiert nach Erregung und Valenz, enthält [28][29]. Nach einer 15 minütigen Entspannungsphase legt man den Versuchspersonen jeweils 15 Sekunden lang eine Auswahl aus diesen Fotos vor. Für die Erstellung der Datenbank des MIT [19][20] wurde eine Person mit schauspielerischer Erfahrung aufgefordert, sich für mehrere Minuten in insgesamt acht verschiedene emotionale Zustände zu versetzen. Als Hilfe rief sie sich passende Bilder vor Augen, z. B. den Verlust eines wichtigen Menschen, um Trauer zu erzeugen. Nasoz und Lisetti verwendeten für ihre Aufzeichnungen dagegen kurze Videoclips oder Filmausschnitte [30]. Ein generelles Problem bei der Erstellung eines Datenkorpus ist, dass eine Versuchsperson über einen längeren Zeitraum stets in ähnlichen Situationen aufgezeichnet werden muss. In einem Praktikum an der Universität Augsburg, in dem kurze Filme oder Videoclips zur Emotionserregung eingesetzt wurden, zeigte sich jedoch, dass die Versuchspersonen auf Medien, die sie bereits gesehen hatten, nur noch schwach oder gar nicht mehr reagierten. Für das vorliegende Experiment sollte deshalb Musik helfen sich in die gewünschten Emotionen zu versetzen. Im Gegensatz zu Bildern oder Filmen, hat Musik den Vorteil, dass wir gewöhnt sind, auch über längere Zeiträume immer wieder dieselben Stücke zu hören. Häufig hören wir Musik auch während wir mit anderen Dingen beschäftigt sind, so dass wir Tätigkeiten und Stimmungen mit bestimmten Liedern verknüpfen. Die Erinnerungen, die mit diesen Liedern verbunden sind, können uns helfen, sehr schnell wieder in diese Stimmung zu wechseln. Natürlich sind diese Assoziationen individuell, da sie auf persönlichen Erlebnissen basieren. Auch die Einschätzung welche Lieder wir z. B. als angenehm oder störend empfinden hängt stark von unserem persönlichen Musikgeschmack ab und lässt sich nur schwer verallgemeinern. Die Entwickler des „Affective DJ“ (vgl. 3.2.3) stellten fest, dass sich mit den Musikstücken, die sie herausgesucht hatten, um ihre Probanden in einen erregenden Zustand zu versetzen, nicht immer der gewünschte Effekt erzielen ließ. Bei der anschließenden Befragung zeigte sich dann, dass zwei der vier Versuchspersonen einige der ausgewählten Songs nach einiger Zeit als nervenden empfanden und wohl deshalb nicht wie erwartet darauf reagierten [17]. Um dies zu verhindern, wurden die Versuchspersonen für unser Experiment aufgefordert die Lieder selbst auszuwählen. Darüberhinaus garantiert dies auch einen kontrollierten Ablauf des Experiments, da der Proband vorab weiß, welche Emotion er zeigen

¹International Affective Picture System

soll und die Möglichkeit hat die Musikstücke auszuwählen, die ihn am besten in den gewünschten emotionalen Zustand versetzen.

5.1.4 Reihenfolge

Schließlich sollte auch die Reihenfolge, in der die Emotionen erzeugt werden beachtet sein. Ein stark erregter Organismus braucht einige Zeit, bis er wieder auf ein neutrales Level absinkt. Werden verschiedene Emotionen ohne längere Pausen direkt hintereinander erzeugt, kann es deshalb sinnvoll sein, nach Erregungsleveln zu sortieren und mit den weniger erregenden Beispielen aus den Quadranten III und IV (vgl. 2.3) zu beginnen. Eine andere Möglichkeit ist, die Emotionen gemäß ihrem Anteil an Lust zu ordnen, also solche Beispiele in Quadranten den III und IV bzw. I und II zusammenzufassen. Das Experiment in [27] beginnt daher zunächst mit fünf positiven und wenig erregenden Fotos, zeigt dann fünf negative und wenig erregende Bilder, wechselt dann auf fünf negative und durchschnittlich erregende, anschließend folgen fünf positive bei gleichem Erregungslevel usw. Versetzt sich eine Versuchsperson in eine Emotion, so geschieht dies stets vor dem Hintergrund ihrer tagesabhängigen Stimmung. Man kann davon ausgehen, dass es ihr leichter fällt solche Zustände anzunehmen, die dieser Stimmung entsprechen. In unserem Experiment wurde deshalb bewußt auf eine feste Reihenfolge verzichtet. Stattdessen wurde jeweils die Emotion als erstes gewählt, die der aktuellen Stimmung am nächsten kam. Auch für den weiteren Ablauf gab es keine feste Vorgabe, allerdings sollten die Pausen zwischen den Liedern lang genug sein, damit sich der Organismus wieder auf ein neutrales Level einpendeln kann.

5.2 Beschreibung des Experiments

Das Experiment zur Aufzeichnung des Datenkorpus wurde nun entsprechend der beschriebenen Kriterien entworfen. Die vier Emotionen Freude, Glückseligkeit, Trauer und Ärger wurden aufgezeichnet, wobei jeweils ein selbstgewähltes Musikstück für jede Emotion dem Probanden dabei half, sich in die gewünschte Stimmung zu versetzen. Die in Kapitel 4.2 beschriebenen Signale wurden mit dem ProComp+-Gerät aufgezeichnet. Je nach Länge der Lieder, blieb die Aufzeichnungslänge variabel. Eine Reihenfolge der Lieder war nicht vorgeben, allerdings wurde mit dem Lied begonnen, das am besten der jeweiligen Tagesstimmung entsprach. Die Pausen zwischen den Liedern wurden so lange gewählt, dass sich die Signale wieder auf ein neutrales Level einpendeln konnten. Vor der Aufzeichnung wurde ein Fragebogen mit Fragen bezüglich äußerer Bedingungen (z. B. Uhrzeit, Temperatur, Umgebungsgereusche) und dem persönlichen Tageszustand (z. B. Stimmung, Stressfaktor, Fitness) ausgefüllt. Nach der Aufzeichnung wurde die Versuchsperson außerdem noch

gebeten einzuschätzen, ob es ihr nach eigenem Ermessen gut oder schlecht gelungen war, sich in die jeweilige Emotion hineinzusetzen. Um möglichst konstante Aufnahmebedingungen zu garantieren, wurde ein eigener Raum für die Aufnahmen zur Verfügung gestellt. Abbildung 5.2 zeigt den schematischen Ablauf des Experiments.

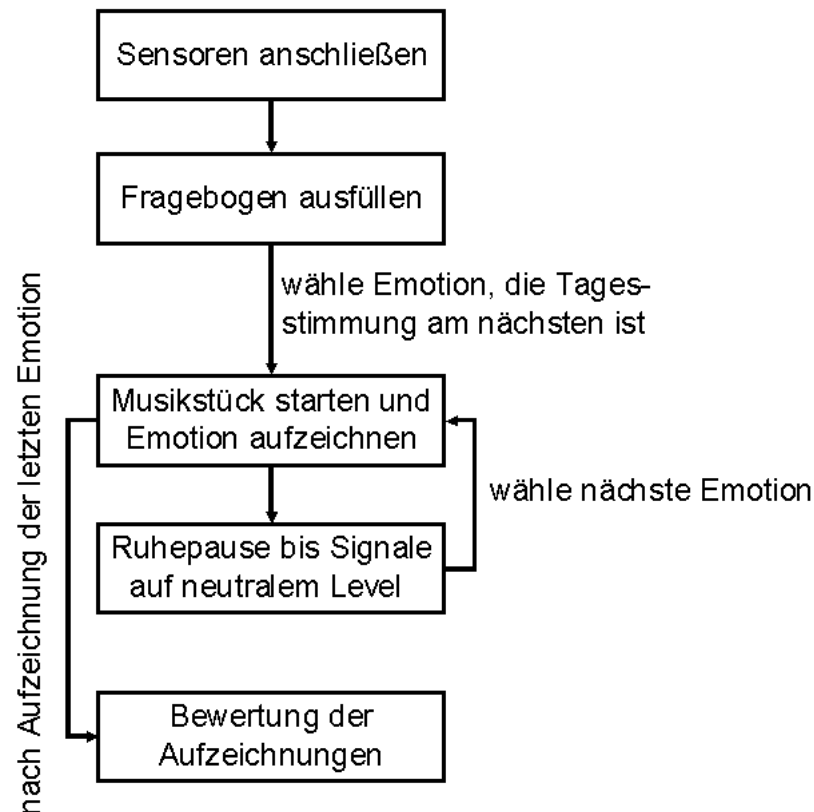


Abbildung 5.2: Schematischer Ablauf des Experiments zur Aufzeichnung des eigenen Korpus.

5.3 Beschreibung des Korpus

Der Korpus, mit dem auch die im Anschluss beschriebenen Verfahren zur Mustererkennung getestet wurden, umfasst 25 Aufnahmen von einer Versuchsperson. Eine Aufnahme setzt sich aus dem Mitschnitt der vier Emotionen zusammen, wobei je Emotion alle vier Signale aufgezeichnet wurden. SC, EMG und RSP mit jeweils 32 Hz, das EKG-Signal mit 256 Hz. Die Signale liegen so vor, wie sie vom Encoder geliefert wurden, d. h. es wurden während der Aufzeichnung weder Filter einge-

setzt, noch in einer anderen Form Artefakte entfernt. Aufnahmen, bei denen es zu Messfehlern oder Ausfällen der Signale kam, wurden komplett entfernt und noch am selben Tag erneut aufgezeichnet. Die Aufzeichnungen von drei Tagen wurde jedoch später nicht ausgewertet, da es während der Aufnahme zu Ausfällen des EMG-Sensors gekommen war, die erst im nachhinein aufgefallen waren. Da die Lieder für die jeweiligen Emotionen unterschiedlich Länge hatten, variiert auch die Länge der Signaldateien. Für die Auswertung der Daten wurden die Signale später jedoch auf eine feste Länge zugeschnitten. Wie es der Ablauf des Experiments vorsah, füllte die Versuchsperson jeden Tag einen Fragebogen aus. Die jeweiligen Angaben zur Tagesstimmung wurden später mit dem Klassifikationsergebnis der jeweils zugehörigen Emotion verglichen. Das erwartete Ergebnis, dass diese Emotion dann besonders gut erkannt wurde, konnte jedoch nicht mit Sicherheit bestätigt werden. Tatsächlich lag zwar die Erkennungsquote für Glückseligkeit mit 85% bzw. Freude mit 80% an Tagen, an denen die Person diese Emotion als Tagesstimmung angab höher als der Durchschnitt (73% für Glückseligkeit bzw. 77% für Freude, siehe Kapitel 7.1.1), allerdings wurden an den vier Tagen, an denen der Proband Trauer als Grundstimmung angab, nur an einem Tag Trauer richtig erkannt. Ärger als Grundstimmung wurde zudem keinmal angegeben. Auf eine weitere Auswertung der Fragebögen wurde deshalb zunächst verzichtet bis mehr Daten vorliegen.

Kapitel 6

Vom Signal zur Emotion

Dieses Kapitel befasst sich mit den nötigen Schritten zur Entwicklung eines Systems, das in der Lage ist Emotionen aus physiologischen Daten zu erkennen. Abbildung 6.1 zeigt den schematischen Ablauf der Mustererkennung. Nach Aufbau eines Korpus werden zunächst für jedes Signal eine Reihe von Features berechnet. Im nächsten Schritt versucht man dann verrauschte und redundante Information aus dem Feature-Set zu entfernen. Das Ergebnis der Reduktion dient danach als Input für den Aufbau des Klassifikators. In der Evaluationsphase wird dieser dann getestet. Im Folgenden können dann einzelne Schritte nochmals verbessert werden, bis schließlich die gewünschte Klassifikationsquote erreicht wird. Alle vorgestellten Verfahren wurden in MATLAB [32] implementiert und mit dem eigenen Datenkorpus getestet (vgl. Kapitel 5). Die erzielten Ergebnisse sind in Kapitel 7 zusammengefasst.

6.1 Berechnung relevanter Features

Für die Klassifikation benutzt man nicht die Werte des rohen Signals, sondern berechnet stattdessen eine Reihe von Kennwerten. Diese sog. Features erfassen die charakteristischen Eigenschaften des Signals und werden als Vektor gespeichert. Dies reduziert die Datenmenge und steigert gleichzeitig die Performanz des Klassifikators. In der Literatur wird die Berechnung der Kennwerte manchmal auch als Feature-Extraktion bezeichnet. Dieser Begriff wird hier mit Bedacht vermieden, da er gelegentlich auch für die, im Anschluss beschriebene Reduktion des Feature-Raums, verwendet wird. Vor Berechnung der Features wurden die Signale zunächst auf zwei Minuten gekürzt und die letzten drei Aufzeichnungen aufgrund eines Ausfalls des EMG-Sensors entfernt. Die Fensterlänge legt die Größe des Intervalls fest, für das jeweils ein Featurevektor berechnet wird. Für die folgenden Untersuchungen wurde je Aufnahmetag und Emotion genau ein Featurevektor berechnet,

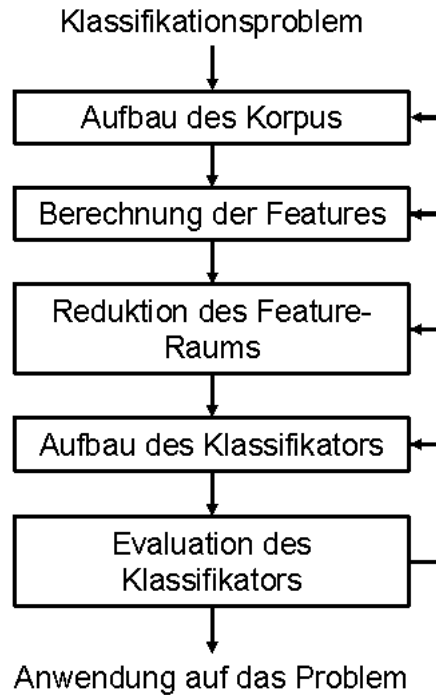


Abbildung 6.1: Schematischer Ablauf der Mustererkennung.

d. h. die Fensterlänge der Länge einer Aufzeichnung angepasst. Die Signal- und Fensterlänge entspricht damit der, die auch Healey in [19] verwendet, und ermöglicht somit einen Vergleich beider Korpora in Abschnitt 7.3.

6.1.1 Basisfeatures

Zunächst wurden für das geglättete und normalisierte SC- bzw. EMG-Signal und für Atem- bzw. Herzfrequenz, die aus dem RSP- bzw. EKG-Signal ermittelt wurden, eine Reihe von Basisfeatures berechnet. Bezeichne X_n jeweils den n -ten Samplewert eines Signals der Länge N , so berechnen diese sich wie folgt:

1. Mittelwert:

$$\mu = \frac{1}{N} \sum_{n=1}^N X_n \quad (6.1)$$

2. Varianz:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (X_n - \mu_X)^2} \quad (6.2)$$

3. Mittelwert der Ableitung (Absolutwert der ersten Differenz):

$$\delta = \frac{1}{N} \sum_{n=1}^{N-1} |X_{n+1} - X_{n-1}|. \quad (6.3)$$

Die Glättung des SC- bzw. EMG-Signals erfolgte mit einem Lowpassfilter und einer Cutoff-Frequenz von 0.3 bzw. 1.0 Hz. Die Normalisierung mit Hilfe folgender Formel:

$$\tilde{X}_n = \frac{X_n - \mu_X}{\sigma_X}. \quad (6.4)$$

Aus dem geglätteten RSP-Signal wurden außerdem noch vier Frequenzbänder im Bereich von 0-0.4 Hz bestimmt. Für deren Berechnung sei an dieser Stelle auf [19] verwiesen.

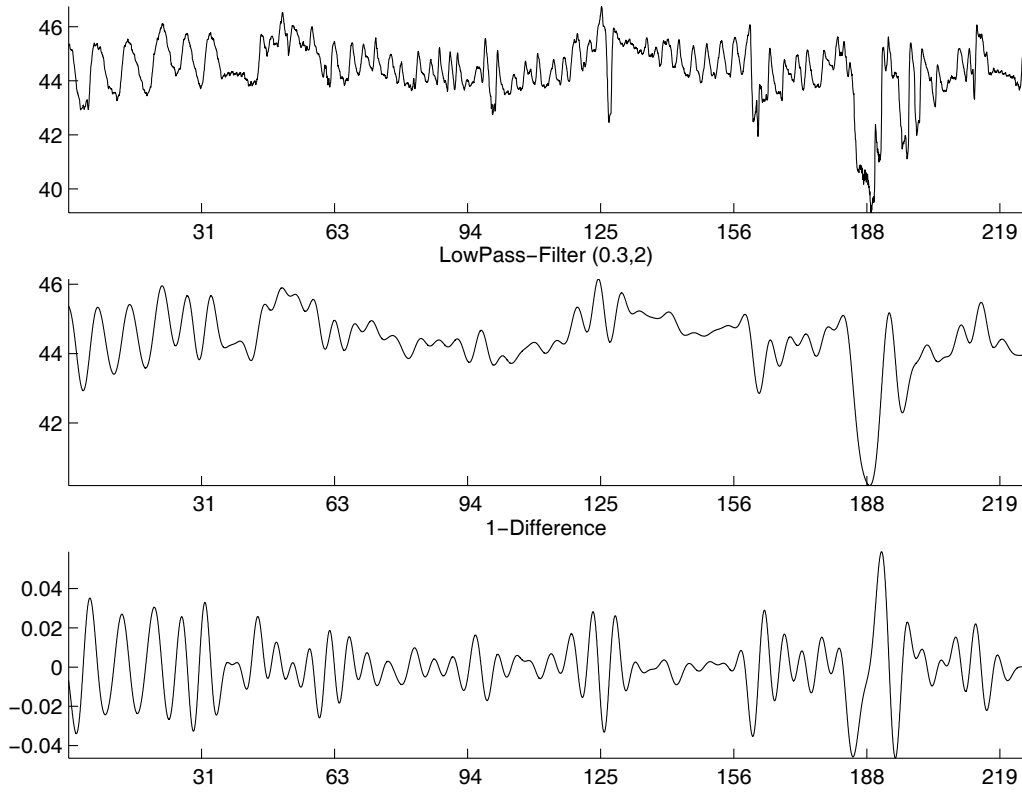


Abbildung 6.2: Damit die Atemfrequenz berechnet werden kann, wird das rohe RSP-Signal (o.) zunächst mit einem Lowpassfilter (m.) geglättet. Nach Berechnung der Ableitung (u.) erhält man die gewünschte Schwingung um die x-Achse.

6.1.2 Berechnung der Atemfrequenz

Das rohe RSP-Signal ist geprägt durch Artefakte. Häufige Abweichungen von der Grundlinie machen es unmöglich die Frequenz direkt zu bestimmen. Zunächst wurde das Signal deshalb mit einem Lowpassfilter geglättet. Berechnet man anschließend die Ableitung, so erhält man die gewünschte regelmäßige Schwingung um die x-Achse. Aus dem Abstand zweier aufeinanderfolgender Minima und Maxima kann dann die Atmungsfrequenz berechnet werden (Abb. 6.2).

6.1.3 Berechnung der Herzfrequenz

Die Herzfrequenz kann aus dem Abstand aufeinanderfolgender R-Zacken im EKG-Signal bestimmt werden. Dazu sucht man nach Maximas im Frequenzspektrum, das im Bereich der R-Zacke am dichtesten ist. Abb. 6.3 zeigt ein EKG-Signal, das von seinem Power-Spektrum überlagert wird. Man sieht wie die R-Zacken des rohen Signals und die Maximas im Power-Spektrum zusammenfallen.

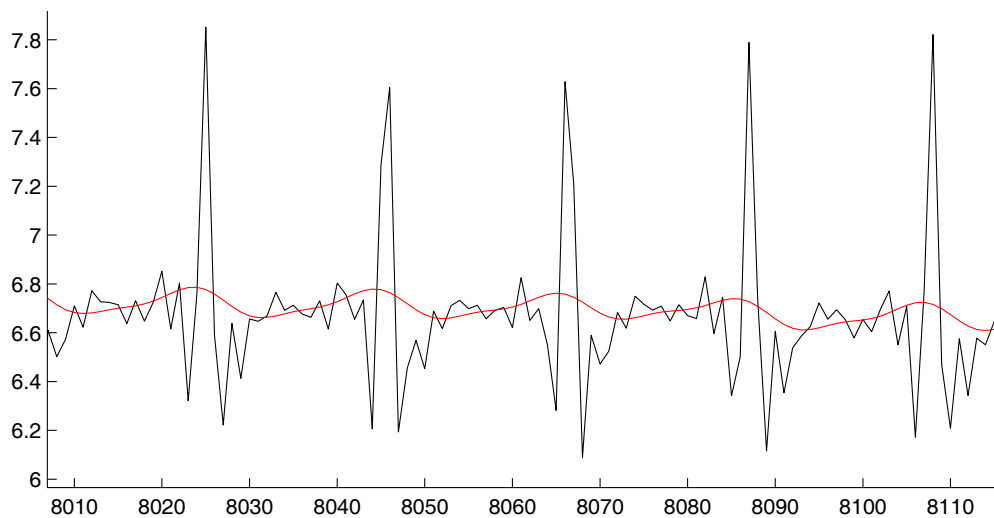


Abbildung 6.3: EKG (schwarz) mit überlagertem Power-Spektrum (rot).

6.1.4 Zusätzliche Features

Detektion der SC-Grundlinie

Der SC-Sensor misst die Schweißproduktion an den Fingerkuppen. Diese steigt z. B. wenn wir unter Stress geraten und fällt wieder bei Entspannung. Man kann dies deutlich an den Ausschlägen im Signal erkennen. Die Amplitude eines Ausschlags ergibt sich aus dem maximalen Wert relativ zur Grundlinie und ist ein wichtiger

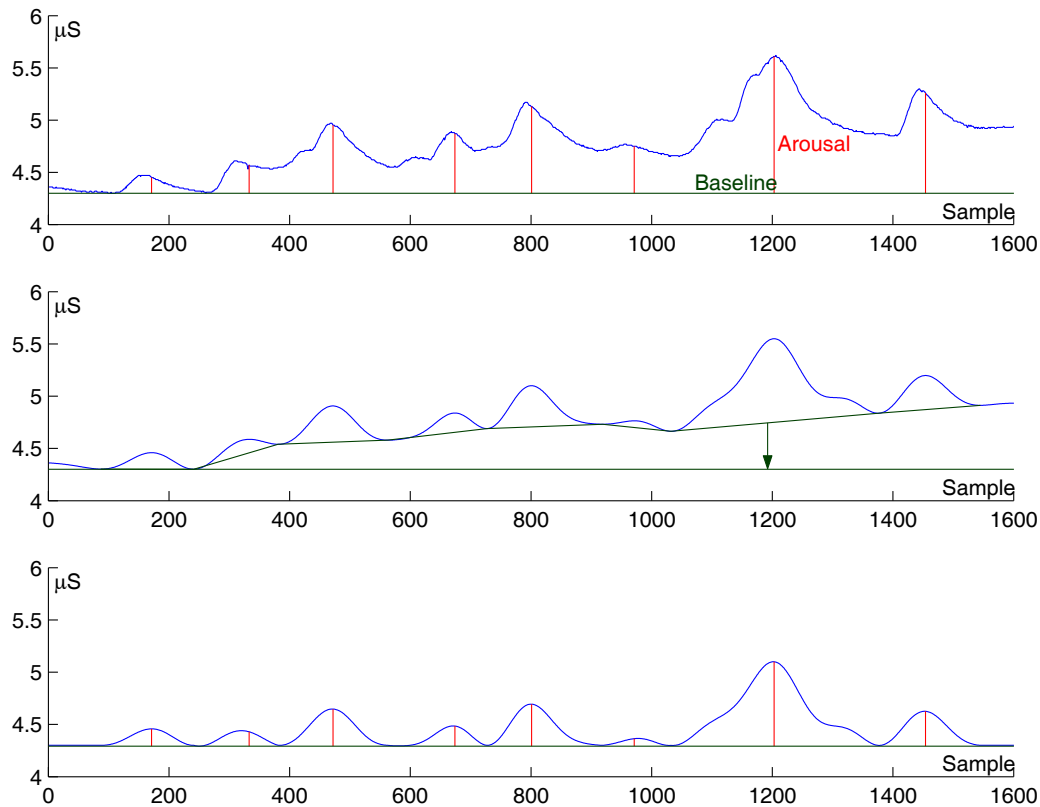


Abbildung 6.4: Der erste Graph zeigt das Rohsignal. Im zweiten Bild wurde das Signal geglättet und die Grundlinie bestimmt. Die dritte Abbildung zeigt das Ergebnis nach Anpassen des Basislevels.

Kennwert für den Hautleitwert. Bei Messungen über mehrere Minuten kann das Basislevel des Signals unter Umständen stark schwanken, denn der Hautleitwert erholt sich nur sehr langsam, so dass weitere Ausschläge bereits erfolgen können, bevor er auf den alten Wert zurückgefallen ist. Misst man die neue Amplitude zur alten Grundlinie, erhält man einen wesentlich höheren Ausschlag als dies eigentlich der Fall ist. Messfehler dieser Art können nur verhindert werden, indem das Level des Signals nach jedem Ausschlag angepasst wird. Abbildung 6.4 zeigt ein Beispiel für die Anpassung der Grundlinie. Das rohe Signal wird zunächst mit einem Lowpassfilter und einer Cutoff-Frequenz von 0.3 Hz geglättet. Anschließend werden alle Minimas berechnet und zu einer Linie verbunden. Diese Linie ist eine Annäherung der Grundlinie und wird von dem Signal subtrahiert. Die grafische Analyse mit Hilfe eines Boxplots (siehe Kapitel 6.1.5) in Abbildung 6.5 vergleicht die neue Verteilung der Featurewerte vor und nach Anwendung des Algorithmus.

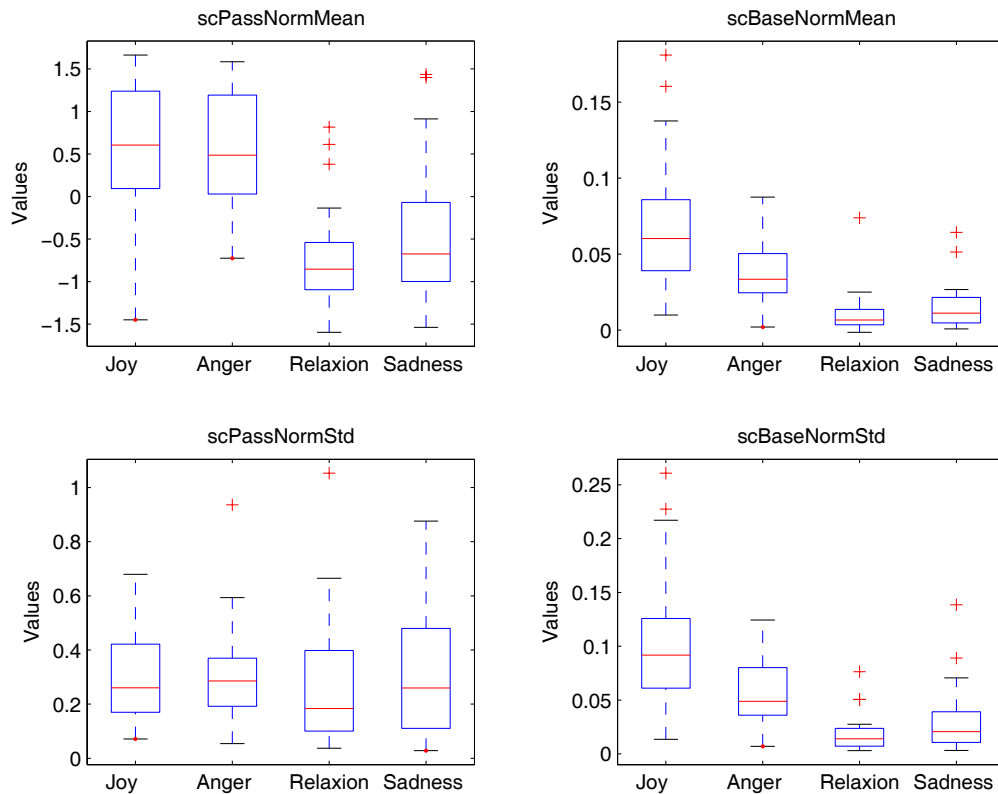


Abbildung 6.5: Boxplot von Mittelwert (o.) und Varianz (u.) des rohen (l.) und auf die Grundlinie angepassten (r.) SC-Signals.

EMG ohne Artefakte

Das EMG dient der Erfassung der elektrischen Muskelaktivität. Da Muskeln sehr empfindlich auf Bewegung reagieren, können sich schnell ungewollte Artefakte im Signal niederschlagen. Zudem hat das EMG-Signal ein sehr breites Frequenzspektrum, das die Anwendung von Hochpass- und Bandpassfiltern teilweise ausschließt. Allein das Heben und Senken des Brustkorbs kann schon ausreichen, um das Signal zu stören. Bisweilen lässt es sich auch nicht verhindern, dass ein EKG-Signal mitregistriert wird, insbesondere bei Ableitungen von der Rumpfmuskulatur [25]. Auch in den eigenen Aufzeichnungen, bei denen die Elektroden im Nacken der Person befestigt wurden, fanden sich beide Artefaktquellen. Um die durch die Atmung bedingte Störung zu entfernen, genügt ein Lowpassfilter mit einer Cutoff-Frequenz von 1.0 Hz. Das geglättete Signal weist eine erstaunlich starke Korrelation mit dem RSP auf und wird einfach subtrahiert. Mit Hilfe der R-Zacken aus dem parallel aufgezeichneten EKG-Signal lassen sich außerdem die Bereiche der zweiten Störung lokalisieren. Es genügt dort das Maximum durch den Wert des rechten oder linken

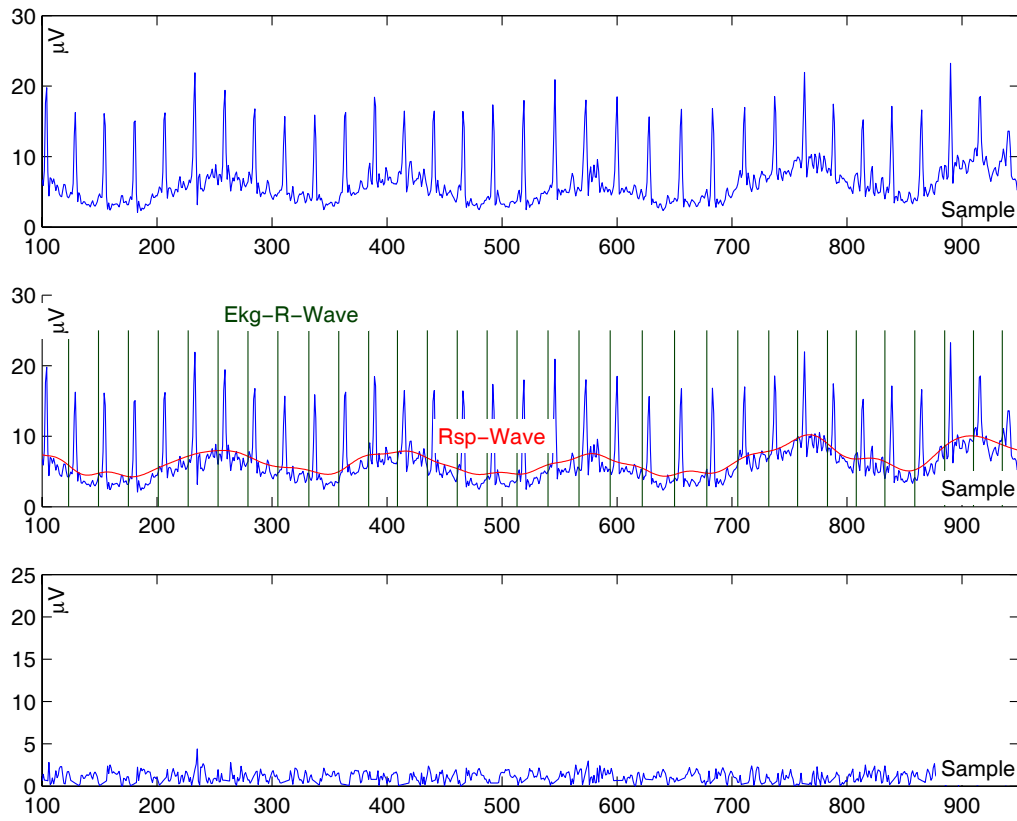


Abbildung 6.6: Der erste Graph zeigt noch das, von Artefakten geprägte Rohsignal. Im zweiten Bild ist der Zusammenhang der Störung mit dem EKG-Signal und der Atemfrequenz sichtbar gemacht. In der letzten Abbildung sieht man das Signal nach Entfernen der Artefakte.

Nachbarn zu ersetzen. Abbildung 6.6 zeigt das Signal vor und nach Anwendung des Algorithmus. Die Artefakte im EMG zeigen aber auch, dass vermeintliche Störungen interessante Informationen über andere Signale enthalten können. Würde es gelingen Atem- und Herzfrequenz zuverlässig aus dem EMG-Signal zu berechnen, müsste man diese Signale nicht noch extra aufzeichnen. Weniger Sensoren bedeuten aber weniger Aufwand und Einschränkung für den Benutzer.

RSP-Amplitude

Mit Hilfe eines elastischen Gurts, der bei Brustatmern um die Brust und bei Bauchatmern um den Bauch gelegt wird, wird die Atmung des Benutzers gemessen. Neben der Atemfrequenz kann auch die Amplitude, also die maximale Auslenkung des Bandes, berechnet werden. Diese entspricht, wie in Abbildung 6.7 zu sehen,

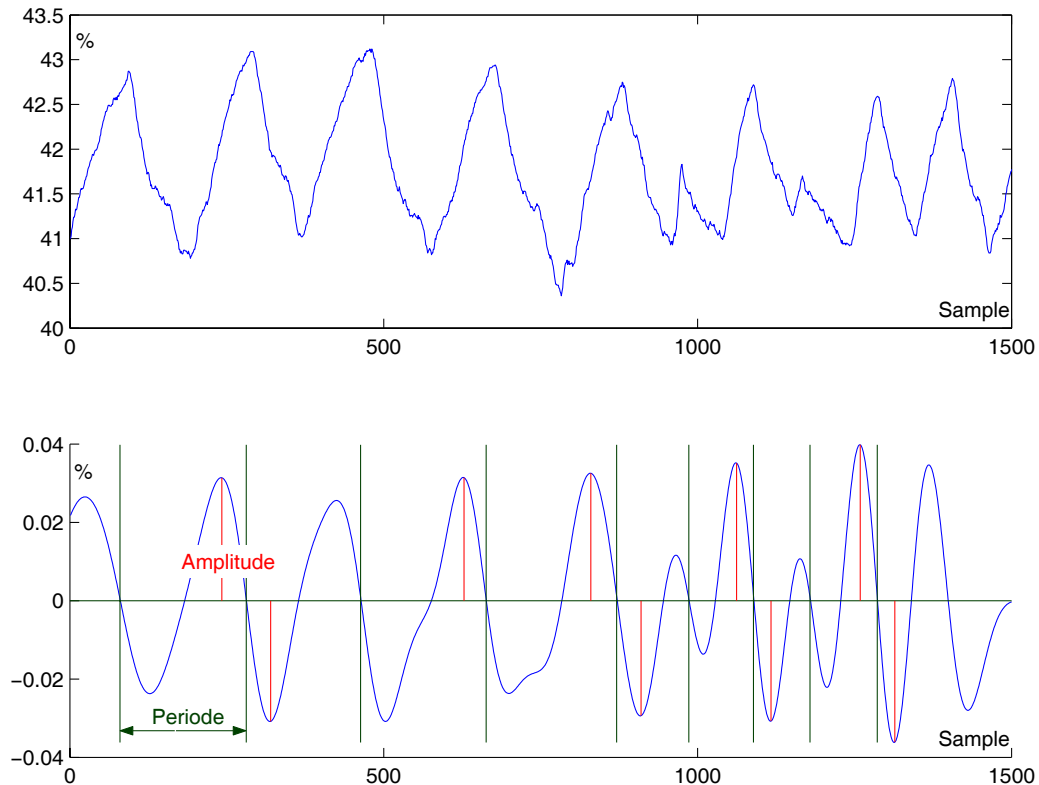


Abbildung 6.7: Der erste Graph zeigt das rohe Atemsignal. Aus dem geglätteten Signal werden Periode und Amplitude bestimmt, wie im Bild darunter zu sehen ist.

dem Maximum der Absolutwerte einer Periode.

6.1.5 Grafische Analyse der Features

Mit Hilfe eines sog. Boxplot kann man die statistische Verteilung der Featurewerte zwischen den Klassen grafisch darstellen. Eine Box grenzt dabei den Bereich ein, in dem die Werte liegen, und wird mit einer horizontalen Linie an der Stelle des Mediums geteilt. Die beiden dadurch entstehenden Teilboxen umfassen je ein Viertel der Werte, die über bzw. unter dem Medium liegen. Die Linien am oberen und unteren Ende der Box, die auch Whiskers genannt werden, begrenzen die Verteilung der restlichen Werte. Ausreißer jenseits der Whiskers werden mit einem \times gekennzeichnet. Der Boxplot vermittelt einen grafischen Eindruck darüber, wie gut sich ein Feature für die Klassifikation eignet. Abbildung 6.8 zeigt den Boxplot der statistischen Verteilung der mittleren Atemfrequenz. Man sieht, dass die Werte dieses Kennwerts bei Ärger deutlich höher liegen als bei den drei anderen Klassen, also

helfen kann, diese Emotion von den anderen Zuständen zu unterscheiden. Signifikanzanalyse ist ein wichtiges statistisches Verfahren zur Bewertung von Features und wird uns auch in den folgenden Kapiteln noch begegnen.

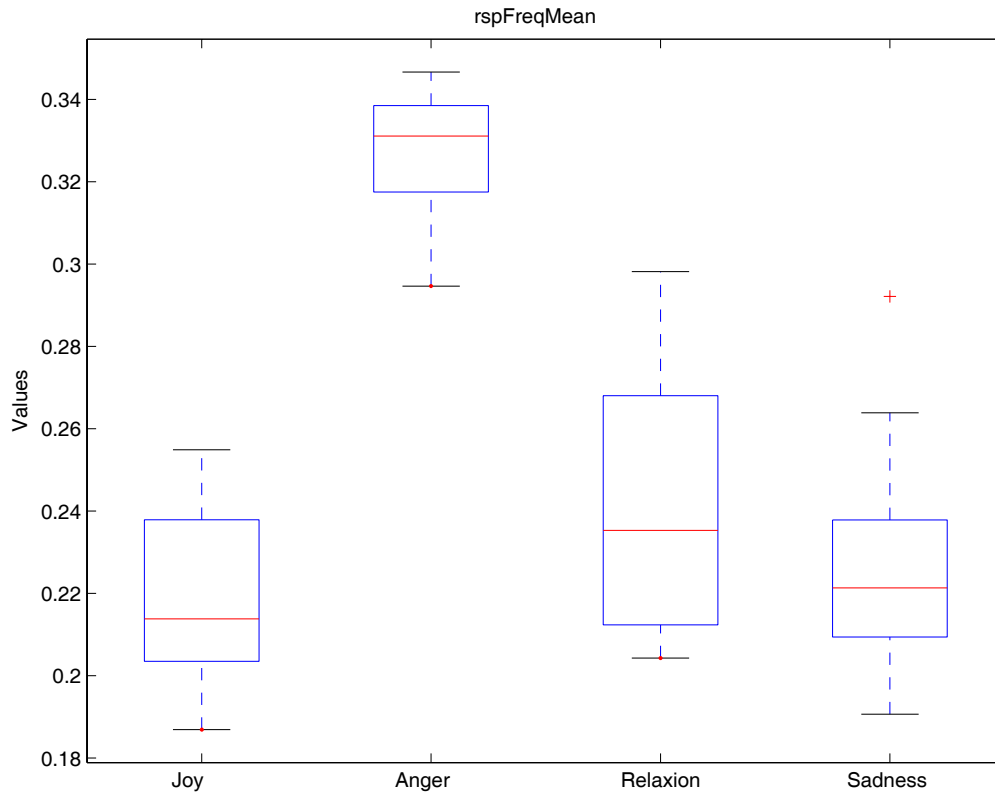


Abbildung 6.8: Boxplot der statistischen Verteilung der mittleren Atemfrequenz zwischen den Klassen.

6.2 Reduktion des Feature-Raums

Hat man eine Reihe von Features für jedes Signal gefunden, versucht man nun den Raum, der von ihnen aufgespannt wird, zunächst wieder zu reduzieren. Besonders zwei Argumente rechtfertigen diesen Aufwand. Zum einen wird die Klassifikation vereinfacht und damit der Rechenaufwand verringert, zum anderen wird, selbst wenn Performance keine Rolle spielen sollte, in der Regel die Effizienz des Klassifikators gesteigert. Letzteres muss zwar nicht zwangsläufig zutreffen, ist jedoch besonders dann wahrscheinlich, wenn einer großen Anzahl von Kennwerten eine relativ kleine Menge an Trainingsdaten gegenüber steht („Curse of Dimensionality“).

[33]). Kandidaten für die Reduktion sind Features, die entweder redundante Information enthalten, z. B. wegen ihrer lineare Abhängigkeit zu anderen, oder zu starkem Rauschen innerhalb des Sets führen und damit die Klassifikation erschweren. Ob durch eine Reduktion des Feature-Raums bessere Ergebnisse zu erwarten sind, hängt aber auch vom Klassifikator selbst ab. Neuronale Netze sind z. B. dafür bekannt, vorausgesetzt der Trainingskorpus ist groß genug, auch mit irrelevanten Daten gut zurecht zu kommen. Bei Entscheidungsbäumen wie C4.5 ist dagegen immer mit einer Reduzierung des Verzweigungsgrads und der Tiefe zu rechnen. Es folgt zunächst ein Überblick über gängige Verfahren zur Dimensionsreduktion. Anschließend wird auf die, in dieser Arbeit angewandten Strategien und die resultierenden Ergebnisse eingegangen.

6.2.1 Überblick über gängige Methoden

Das Ziel ist, ein Feature-Set zu finden, das kleiner als das gegebene ist und einen möglichst geringen Klassifikationsfehler bedingt, idealerweise natürlich einen kleineren als das alte Set. Wie nehmen an, die Dimension eines gegebenen Feature-Raums der Größe d soll auf m verringert werden ($m < d$). Dies kann entweder erreicht werden, indem $m - d$ Features komplett aus dem Set entfernt werden, oder mit Hilfe einer Projektion, die die gegebenen Features in einen m -dimensionalen Raum abbildet. Die Suche nach einem Subset der Größe m kann entweder erschöpfend, also eine Betrachtung aller Möglichkeiten, entsprechend einer Heuristik oder durch zufällige Auswahl erfolgen [38].

Selektionsverfahren

Der naive Ansatz für ein Selektionsverfahren besteht darin, alle möglichen Kombinationen von d Features zu testen und mit der besten weiterzuarbeiten („Exhaustive Search“). Dafür müssen dann exakt $\binom{d}{m}$ Fälle überprüft werden. Selbst wenn die Selektion offline erfolgt, ist der Aufwand für größere Feature-Sets nicht tragbar. Aus diesem Grund wurden Algorithmen entwickelt, die entsprechend einer Heuristik nur noch einen Teil der Möglichkeiten betrachten, indem sie ausgehend von bereits als gut bewerteten Subsets schrittweise Feature hinzunehmen oder entfernen. Die SFS-Methode (Sequential Forward Selection) startet z. B. mit dem besten individuellen Feature und nimmt dann in jedem Schritt jeweils das Feature hinzu, welches das Set am besten ergänzt. Beim SBS (Sequential Backward Search) geht man dagegen von dem vollen Set aus und entfernt in jedem Schritt das schlechteste. Variationen wie der SFFS-Algorithmus (Sequential Forward Floating Search) oder die „Plus l-take away r“-Methode versuchen durch eine Kombination von Vor- und Rückwärtssuche die Gefahr, in einem lokalen Maximum steckenzubleiben, zu verringern. Ein optimales Feature-Set können aber auch sie nicht garantieren.

Projektionsverfahren

Statt Features ganz aus dem Set zu entfernen, und damit unter Umständen diskriminante Information zu verlieren, kann die gesamte Information des alten Feature-Raums dazu benutzt werden, um daraus neue Kennwerte für ein Subset zu berechnen. Da es keinen systematischen Weg gibt, mit der sich eine nichtlineare Funktion $y = f(x)$ finden lässt, die das Feature-Set x optimal in ein neues Set y mit geringerer Dimension überführt, greift man meist auf eine lineare Projektion W zurück. Die Abbildung sollte dabei so gewählt werden, dass möglichst viel relevante Information erhalten bleibt. Der Vorteil gegenüber dem Selektionsverfahren ist, dass verrauschte Features nicht komplett entfernt werden, sondern ihre signifikanten Anteile erhalten bleiben. Da keine Kombinationen verschiedener Feature-Sets geprüft werden müssen, ist die Berechnungszeit in der Regel kürzer, dafür müssen aber auch weiterhin alle Features berechnet werden. Gängige Methoden sind PCA (Principal Component Analysis) oder die Fisher Transformation. Auch Lösungen mit Neuronalen Netzen sind bekannt (z. B. PCA Network oder Self-Organizing Map (SOM)) [33].

Evaluationsfunktionen

Die Bewertung eines reduzierten Sets geschieht mit einer Evaluationsfunktion $J(\cdot)$. Ein „wrapper“ greift dafür auf den Klassifikator selbst zurück. Die Autoren von [39] bezeichnen den Klassifikationsfehler als den „einzigen aussichtsreichen und legitimen“ Maßstab zur Bewertung eines Subsets. Der Einsatz von „Filtern“ zur Evaluation kann jedoch viel Rechenzeit sparen. Filter bewerten die Features z. B. anhand ihres Abstands innerhalb und zwischen den Klassen, ihres Informationsgehalts oder messen die Korrelation zwischen ihnen [38].

6.2.2 In dieser Arbeit eingesetzte Verfahren

Auch wenn es keine allgemeingültigen Regeln für den Zusammenhang zwischen der Featurezahl f und Trainingsdaten t in Bezug zum Klassifikationsfehler gibt, wird in der Literatur ein Verhältnis von $f/t > 10$ vorgeschlagen [37]. Die Tatsache, dass dieses Verhältnis wegen der relativ geringen Anzahl an Trainingsdaten, die für diese Arbeit zu Verfügung standen, nicht erreicht wird, spricht für den Einsatz dimensionsreduzierender Maßnahmen. Sowohl Selektions- als auch Projektionsverfahren wurden deshalb getestet und miteinander verglichen.

Varianzanalyse

Bei der einfaktoriellen Varianzanalyse (one-way ANOVA) wird jedes Feature darauf untersucht, ob sich seine Werte zwischen den Klassen signifikant unterscheiden

[34][35]. Ein Feature wird bezüglich zweier Klassen als signifikant bezeichnet wenn die Nullhypothese $H_0 : \mu_1 = \mu_2$ nicht zutrifft. Ob die Nullhypothese erfüllt ist, wird mit dem F-Test geprüft. Dabei wird die Varianz innerhalb der Klassen mit der Varianz zwischen den Klassen verglichen:

$$F = \frac{MS_B}{MS_W}. \quad (6.5)$$

Um MS_B und MS_W zu berechnen, bestimmt man zunächst die Gesamtquadratsumme SS_t und die Regressionsquadratsumme SS_b . Sei N die gesamte Anzahl an Samples, C die Anzahl der Klassen und N_c die Anzahl der Samples in Klasse c , und bezeichne x_c^n das n -te Sample der c -ten Klasse, so gilt:

$$SS_t = \sum_{c=1}^C \sum_{n=1}^{N_c} (x_c^n - \mu)^2 \quad (6.6)$$

$$SS_b = \sum_{c=1}^C N_c (\mu_c - \mu)^2. \quad (6.7)$$

Dabei steht μ für das globale Mittel und μ_c für das Mittel der Klasse c . Die Residuenquadratsumme SS_w ist dann:

$$SS_w = SS_t - SS_b. \quad (6.8)$$

Die Quadratsummen dividiert man nun noch durch den entsprechenden Freiheitsgrad df :

$$MS_b = \frac{SS_b}{df_b} = \frac{SS_b}{C - 1} \quad (6.9)$$

$$MS_w = \frac{SS_w}{df_w} = \frac{SS_b}{N - C}. \quad (6.10)$$

Mit der Dichtefunktion der F-Verteilung wird schließlich die Wahrscheinlichkeit berechnet, mit der eine Zufallsgröße in das Intervall $[0..F]$ fällt:

$$f(x|m;n) = m^{\frac{m}{2}} n^{\frac{n}{2}} \cdot \frac{\Gamma(\frac{m}{2} + \frac{n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \frac{x^{\frac{m}{2}-1}}{(mx + n)^{\frac{m+n}{2}}}, \quad (6.11)$$

wobei $\Gamma(x)$ den Wert der Gammafunktion an der Stelle x bezeichnet. Die Wahrscheinlichkeit für F ist dann:

$$p = 1 - f(F|df_b; df_s). \quad (6.12)$$

Bei geringer Varianz zwischen den Klassen liegt p nahe bei 1 und die Nullhypothese wird bestätigt. Ein Ergebnis in der Nähe von 0 bedeutet dagegen, dass die Nullhypothese nicht zutrifft. Der Grenzwert, ab dem die Nullhypothese als erfüllt gilt, heißt Signifikanzlevel. Bei der Feature-Selektion sortiert man die Features entsprechend ihres p -Werts in aufsteigender Reihenfolge und wählt dann die ersten d Features. Abbildung 6.9 zeigt den Klassifikationsfehler in Abhängigkeit von d .

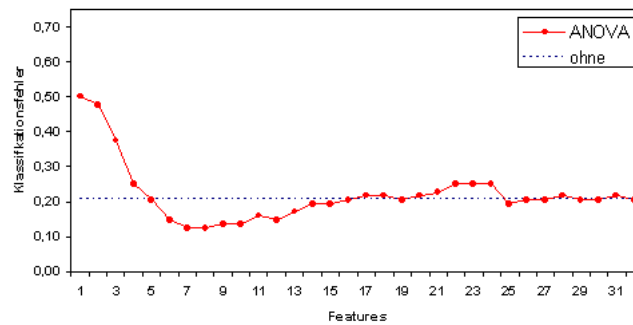


Abbildung 6.9: Verlauf des Klassifikationsfehler mit ANOVA in Abhängigkeit von der Featurezahl.

Einseitige sequentielle Suche: SFS und SBS

Als Selektionsverfahren wurden SFS (Sequential Forward Selection), SBS (Sequential Backward Selection) und SFFS (Sequential Floating Forward Selection) wie in [39] beschrieben implementiert. SFS beginnt mit einem leeren Set und nimmt in jedem Schritt das Feature hinzu, welches das Set am besten ergänzt („step optimal“-Algorithmus). SBS nimmt dagegen das volle Set und entfernt schrittweise das schlechteste. Beide Algorithmen wurden gewählt, da sie sehr schnell zu einem Ergebnis führen. SFS insbesondere, wenn ein sehr kleines Feature-Set angestrebt wird, SBS falls nur wenige Features entfernt werden sollen. In beiden Fällen besteht jedoch das Risiko, in einem lokalen Maximum hängen-zubleiben. Die Graphen in Abbildung 6.10 vergleichen in Abhängigkeit der Featurezahl beide Methoden in Bezug auf Klassifikationsfehler und Laufzeit. Durch Selektion mit SFS kann der

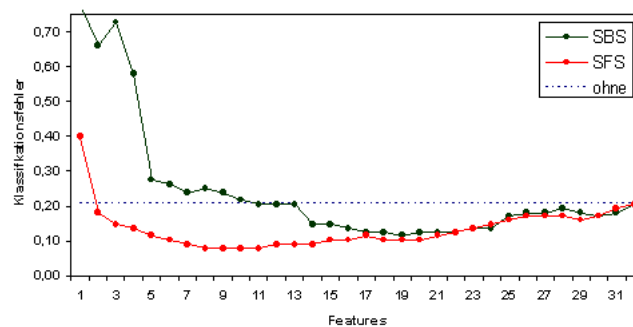


Abbildung 6.10: Verlauf des Klassifikationsfehler mit SFS und SBS in Abhängigkeit von der Featurezahl.

Klassifikationsfehler beinahe durchgehend gesenkt werden. Bei einer Reduzierung des Sets von 32 auf 8 - 11 Features um immerhin 12,5%. Im Vergleich dazu ist bei der SBS-Methode nur bis zu einer Reduzierung auf 22 Features eine Verbesserung um maximal 9,1% zu beobachten.

Dimensionsreduktion mit PCA

PCA bedeutet Principal Components Analysis, auch Hauptkomponententransformation genannt und ist ein Projektionsverfahren zur Dimensionsreduktion mehrdimensionaler Datenfelder. PCA nutzt dabei eine lineare Transformation, die ein Maximum an Varianz innerhalb der Daten erhält. Ziel der Hauptachsentransformation ist es, eine Abbildung W zu bestimmen, die beim Übergang in einen dimensional kleineren Raum möglichst viel Information erhält. Das Verfahren wird in der Praxis auch häufig zur Visualisierung von mehrdimensionalen Datenfeldern eingesetzt. Um W zu bestimmen sucht man für die Kovarianzmatrix Σ der Feature-Matrix die folgende Zerlegung:

$$\Sigma = A\Lambda A^T. \quad (6.13)$$

Die Spalten von A sind dabei die orthonormalen Eigenvektoren von Σ , und Λ ist eine Diagonalmatrix mit den Eigenwerten von Σ , $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. Die Eigenvektoren, die zu den m größten Eigenwerten korrespondieren, bilden die Abbildungsmatrix W . Abbildung 6.11 zeigt den Verlauf des Klassifikationsfehlers nach Reduktion des Feature-Raums. Die Grafik macht deutlich, dass die Dimen-

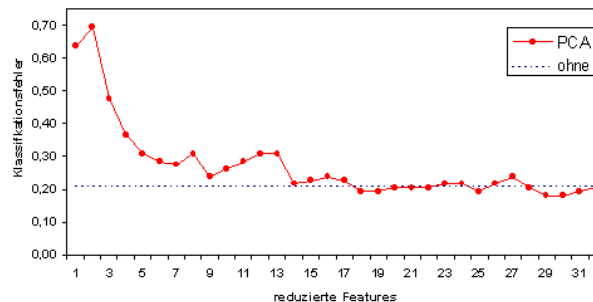


Abbildung 6.11: Verlauf des Klassifikationsfehlers nach Reduktion des Feature-Raums mit PCA.

sionsreduktion mit PCA den Klassifikationsfehler anwachsen lässt, statt ihn zu verbessern. Man stößt hier auf ein typisches Problem, das immer dann auftritt, wenn die Auswahl der Features anhand einer Menge von inhomogenen Trainingsbeispielen erfolgt, ohne Wissen über die Klassenzuordnung zu berücksichtigen. Die

Anwendung des PCA-Verfahrens in der oben beschriebenen Form garantiert nämlich nicht, dass die, für die Klassifizierung wichtigen Informationen erhalten bleiben. Dies ist nur der Fall, wenn die Varianz zwischen und innerhalb der Klassen in derselben Richtung liegen oder die Varianz zwischen den Klassen deutlich größer ist als die innere. Ansonsten führt PCA zu einem teilweisen oder vollständigen Verlust der Diskriminanz [45]. Mit der Principal Feature Analysis (PFA) [44] oder Oriented Principal Component Analysis (OPCA) [45] wurden deshalb Varianten entwickelt, die Informationen über die Klassen in Verbindung mit PCA nutzen.

Fisher-Projektion

Der Erfolg der Klassifikation hängt insbesondere davon ab, wie nah die Muster innerhalb derselben Klasse im Feature-Raum beieinander liegen. Gleichzeitig sollte der Abstand zwischen den Klassen möglichst groß sein. Die Fisher-Projektion, ein Verfahren aus der linearen Diskrimanzanalyse (LDA), nutzt Linearkombinationen der Features, um eine Abbildung zu finden, die diese beiden Kriterien optimiert [40][41][42]. Dafür wird zum einen die Streuung innerhalb jeder Klasse S_w (within-class scatter) und die Streuung zwischen allen Klassen S_b (between-class scatter) wie folgt berechnet:

$$S_W = \frac{1}{N} \sum_{c=1}^C \sum_{n=1}^{N_c} (x_c^n - \mu_c)^t (x_c^n - \mu_c) \quad (6.14)$$

$$S_B = \frac{1}{N} \sum_{c=1}^C N_c (\mu - \mu_c)^t (\mu - \mu_c). \quad (6.15)$$

Hier steht C für die Anzahl der Klassen, N_c für die Anzahl der zugehörigen Samples und N für die Summe aller Samples; μ_c ist der Mittelwertvektor der Klasse c und μ das globale Mittel. Das n -te Sample der Klasse c wird mit x_c^n bezeichnet. Abbildung 6.12 zeigt Beispiele für eine schlechte und eine gute Projektion, und welche Auswirkung dies jeweils auf die Streuung innerhalb und zwischen den Klassen hat. Als Maß für die Diskriminanz der Abbildung W wird die folgende Funktion herangezogen:

$$J(W) = \frac{\|W^T S_w W\|}{\|W^T S_b W\|}. \quad (6.16)$$

Mit einer Eigenwertszerlegung bestimmt man dann die Projektion, die eine optimale Trennschärfe bedingt:

$$S_b W = \lambda_w S_w W. \quad (6.17)$$

Die zugehörigen Eigenvektoren mit Eigenwerten ungleich Null sind linear unabhängig und gegenüber der Transformation unveränderlich. Alle anderen Eigenvektoren

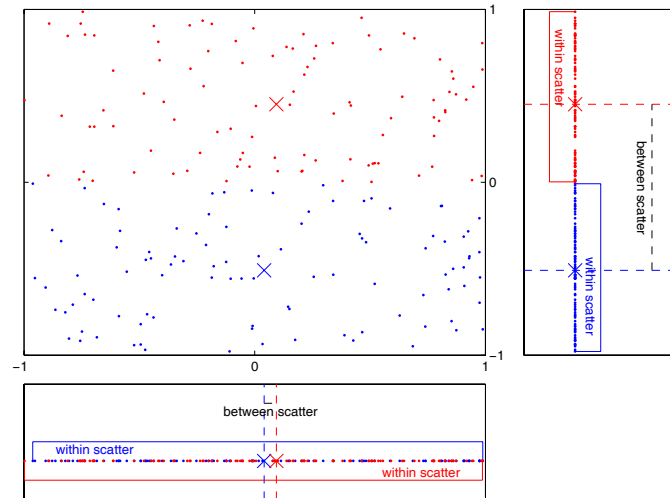


Abbildung 6.12: Die Figur zeigt zwei 2-dimensionale Punktwolken, die in einen eindimensionalen Raum projiziert werden sollen. Die Projektion auf die x-Achse hat zur Folge, dass die beiden Klassen nicht mehr voneinander getrennt werden können, da ihre Zentren zu nahe beieinander liegen. Bei der Projektion auf die y-Achse lassen sich die Klassen dagegen noch immer gut unterscheiden.

sind dagegen linear abhängig. Damit das neue Feature-Set möglichst wenig redundante Informationen enthält, bildet man die Transformationsmatrix deshalb nur aus Eigenvektoren, deren Eigenwerte ungleich Null sind. Da die Mittelwerte der Klassen μ_c in die Berechnung der Streumatrizen S_W und S_B eingehen, gibt es dafür genau $C - 1$ Kandidaten. Für die Abbildungsmatrix wählt man dann entweder alle oder eine Auswahl von ihnen, wobei man im letzteren Fall diejenigen bevorzugt, die zu den größten Eigenwerten korrespondieren [41]. Abbildung 6.13 zeigt die Fisher-Projektion angewandt auf ein Problem mit vier Klassen. Deutlich ist zu sehen, dass der Klassifikationsfehler stark ansteigt, sobald die Dimension gleich oder größer der Klassenzahl ist.

Alternative Evaluationsfunktionen

Zur Bewertung der Feature-Sets benötigt man eine Evaluationsfunktion. Häufig greift man dabei auf den Klassifikator selbst zurück. Abhängig davon, wie viele Feature-Sets geprüft werden müssen, kann die Evaluation aber sehr zeitaufwändig werden. So kann es durchaus Sinn machen, eine schnellere Funktion zur Bewertung der Features zu benutzen. Die Güte der Diskriminanz zwischen den Klassen lässt

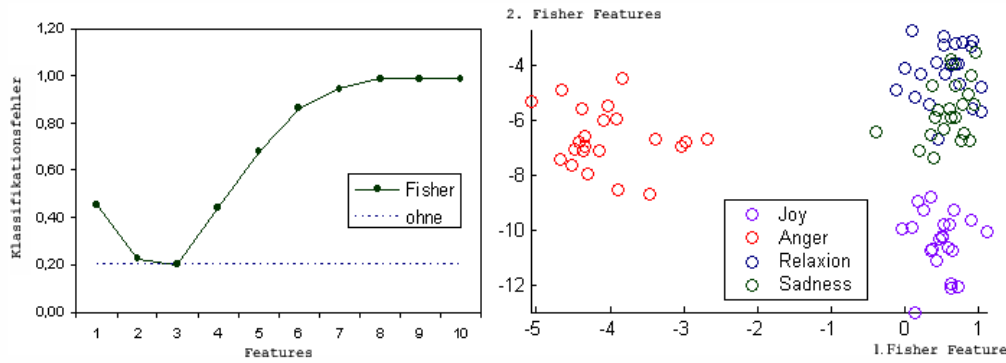


Abbildung 6.13: Verlauf des Klassifikationsfehlers nach Reduktion des Feature-Raums mit der Fisher-Projektion (l.). Häufig wird die Fisher-Projektion auch benutzt, um eine 2-dimensionale Darstellung zu ermöglichen, in der die Klassen möglichst gut voneinander getrennt darstellbar sind (r.).

sich z. B. mit Wilks' Lambda abschätzen [43]:

$$\lambda = \frac{|S_W|}{|S_W + S_B|}. \quad (6.18)$$

Dabei bezeichnet S_W wieder die innere Streumatrix und S_B die Streumatrix zwischen den Klassen. Die Berechnung erfolgt gemäß 6.14 und 6.15. Ein kleiner Wert von λ deutet an, dass sich die Klassen gut unterscheiden lassen. Im optimalen Fall findet eine solche Filterfunktion dasselbe Featureset wie eine Evaluation mit dem Klassifikator selbst.

6.3 Mustererkennung

Nachdem der Feature-Raum, wie im letzten Kapitel beschrieben, reduziert wurde, folgt nun die eigentliche Klassifikation. Dabei versucht man typische Muster zu finden, die eine möglichst gute Trennung der Klassen ermöglichen. Die im Folgenden vorgestellten Verfahren zur Mustererkennung werden deshalb zunächst mit einer Menge von Beispielen trainiert und passen sich damit an das jeweilige Klassifizierungsproblem an. Die Trainingsdaten müssen dafür natürlich in einer vergleichbaren Situation aufgezeichnet werden wie die Daten, die später erkannt werden sollen. Nachdem der Klassifikator trainiert wurde, kann er mit einem weiteren Satz von Daten getestet werden, indem man den Output des Klassifikators mit den tatsächlichen Labels der Testdaten vergleicht. In jedem Fall müssen das Trainings- und Testset disjunkte Mengen sein. Abbildung 6.14 beschreibt den schematischen Ablauf der Klassifikation. Der Datenkorpus wird dabei zunächst in Trainingsdaten,

mit denen der Klassifikator aufgebaut wird, und Testdaten, mit denen anschließend der Klassifikator getestet wird, zerlegt. Wie im letzten Kapitel wird zunächst ein Überblick über gängige Methoden der Mustererkennung gegeben. Anschließend wird auf die Verfahren näher eingegangen, die in dieser Arbeit eingesetzt wurden.

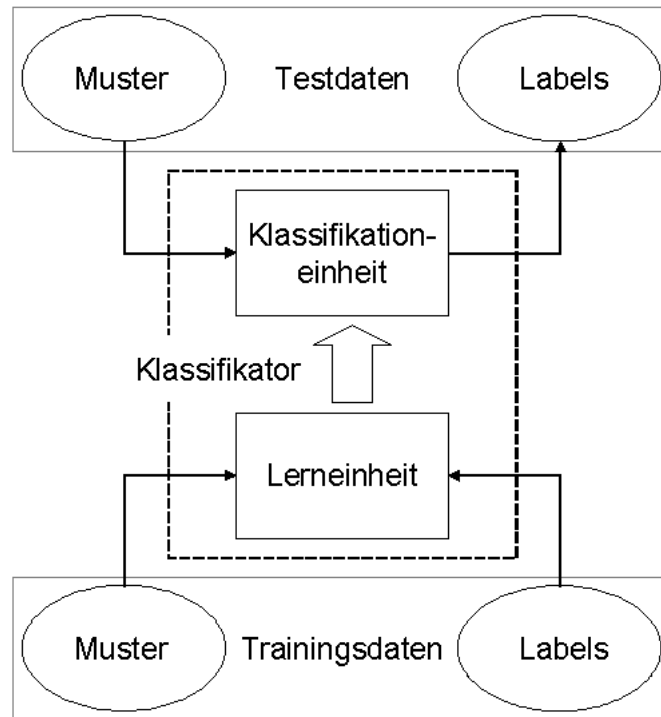


Abbildung 6.14: Schematischer Ablauf der Klassifikation.

6.3.1 Überblick über gängige Methoden

Es gibt eine ganze Reihe verschiedener Ansätze für Klassifikation auf Basis von Mustererkennung [33]. Generell unterscheidet man zwischen überwachtem Lernen (supervised learning) und unüberwachtem Lernen (unsupervised learning). Bei ersterem besteht das Trainingsset aus Paaren von Variablen $\{x, c\}$, wobei x ein Featurevektor und c die zugehörige Klasse ist. In diesem Fall spricht man auch von annotierten oder gelabelten Daten. Unüberwachtes Lernen arbeitet dagegen mit ungelabelten Beispielen. In diesem Fall ist zunächst ein Clustering erforderlich, das ähnliche Beispiele zu gemeinsamen Klassen zusammenfasst. Dieses Problem kann z. B. mit einem Gaussian Mixture Model gelöst werden. Da die Datenkorpora, die für diese Arbeit zur Verfügung standen, bereits annotiert sind, werden im Folgenden nur Strategien für überwachtes Lernen behandelt. Insbesondere wurden

Verfahren getestet, die in anderen Arbeiten bereits erfolgreich bei der Klassifikation von physiologischen Daten eingesetzt wurden.

Ähnlichkeitsanalyse

Der wohl intuitivste Ansatz basiert auf einer Ähnlichkeitsanalyse (Template Matching). Die Samples jeder Klasse werden benutzt, um ein oder mehrere Prototypen zu erstellen. Im einfachsten Fall kann dies z. B. einfach das arithmetische Mittel sein. Die Prototypen werden dann wie eine Schablone gegen das zu klassifizierende Sample gehalten und die Klasse mit der größten Übereinstimmung gewählt. In der Praxis misst man dafür die Distanz oder berechnet die Korrelation.

Statistische Verfahren

Eine weitere Gruppe bilden die statistischen Verfahren. Jedes Sample wird durch n Features repräsentiert und als Punkt in einem n -dimensionalen Raum gesehen. Man sucht nun Features, die in diesem Raum Regionen bilden, die eine Klasse möglichst kompakt zusammenfassen und gleichzeitig von den anderen separieren. In der Regel erstellt man ein Wahrscheinlichkeitsmodell für den Feature-Raum und bemüht sich Grenzen zu finden, die den Raum so aufzuteilen, dass die Klassen möglichst gut voneinander getrennt sind. Dieses Problem kann z. B. mit einer linearen oder quadratischen Diskriminanzfunktion (QDF/LDF) gelöst werden. Dieses Verfahren benutzte auch Healey in [19] zur Klassifikation physiologischer Daten und erreichte für bestimmte Emotionsgruppen Erkennungsquoten von über 80%.

Syntaxanalyse

Für die Syntaxanalyse wählt man eine hierarchische Sicht auf die Pattern, indem man diese in einfache Subpatterns unterteilt. Diese werden wiederum weiter zerlegt, bis man schließlich elementare Pattern, die sog. Primitiven, erhält. Sie stellen das Alphabet dar, die gemäß einer Grammatik wieder zu Sätzen, also den Pattern zusammengesetzt werden können. [46] beschreibt eine Grammatik, die versucht die Wellenform des EKG zu erfassen. Der Satz $[P+, Q, R+, S, T+]$ kennzeichnet z. B. einen normalen Herzschlag mit gut ausgeprägter R-Zacke. Dagegen tritt in der Folge $[P+, Q, R^*, S, T+]$ die R-Zacke nur schwach hervor. Eine solche Grammatik kann helfen, Abnormalitäten im EKG-Signal aufzudecken, um Krankheiten zu erkennen.

Bayessche Netze

Bayessche Netzwerke sind gerichtete azyklische Graphen, die Wahrscheinlichkeitsverteilung verschiedener voneinander unabhängiger Ereignisse modellieren. Ereig-

nisse werden dabei als Knoten repräsentiert. Die Abhängigkeiten zwischen ihnen symbolisieren gerichtete Kanten. Ein Ereignis an der Pfeilspitze hängt dabei stets in irgendeiner Weise von dem am Pfeilende ab. Romain Chabbal beschreibt in [47] ein Bayessches Netzwerk zur Modellierung des emotionalen Zustand eines Benutzers in Abhängigkeit seiner physiologischen Daten.

Neuronale Netze

Die letzte große Gruppe bilden die Neuronalen Netze. Darunter verbergen sich gerichtete Graphen mit gewichteten Kanten, die dem Netz die Fähigkeit geben, komplexe nichtlineare Beziehungen zwischen n Eingängen und m Ausgängen zu erlernen. In der Mustererkennung werden meist mehrstufige Feed Forward Netze (auch Multilayer Perceptron, MLP) oder selbstorganisierende Karten (Selforganizing Map, SOM) eingesetzt. Ein Vorteil der neuronalen Netze ist, dass sie im Gegensatz zu regelbasierten Lösungen vielseitiger einsetzbar sind und sich leicht an neue Problemstellungen anpassen lassen. Künstliche Neuronale Netze, die der Funktionsweise des menschlichen Gehirns nachempfunden sind, gehen auf Untersuchungen von Warren McCulloch und Walter Pitts aus dem Jahr 1943 zurück [49]. Zentraler Baustein dieser Netze ist ein stark idealisiertes Neuron. Ein Neuron wird charakterisiert durch seinen momentanen Aktivierungsgrad und einer Aktivierungsfunktion, die für die Berechnung des neuen Aktivierungsgrads in Abhängigkeit des alten Werts, eines Schwellwerts und einer Netzeingabe zuständig ist. Die Netzeingabe ergibt sich wiederum aus der Ausgabe anderer Neuronen und den Verbindungsgewichten. Eine Ausgabefunktion leitet die Aktivierung des Neurons anschließend an Neuronen der nächsten Schicht weiter. Je nachdem wie sich bei Eingabe der Trainingsmuster der Fehler zwischen erwartetem Wert und der Ausgabe des Netzes verkleinert oder vergrößert, wird der Aktivierungsgrad der einzelnen Neuronen angepasst. Die Verbindung der Neuronen erfolgt durch gerichtete und gewichtete Kanten [50]. Abbildung 6.15 zeigt den schematischen Aufbau eines mehrschichtigen Perzeptrons. In [27] wurde ein neuronales Netz mit 13 Eingängen, einer versteckten Schicht mit zehn Neuronen und einem Ausgangsknoten benutzt, um den Grad an Erregung bzw. Lust aus physiologischen Daten zu erkennen. In den von den Autoren aufgezeichneten Daten, erzielten sie für die beiden Zustände eine Erkennungsquote von 96,6% bzw. 89%.

6.3.2 In dieser Arbeit eingesetzte Verfahren

Allgemein besteht die Aufgabe eines Klassifikators darin, den Raum, der von den Features aufgespannt wird, in Regionen zu zerlegen, die einer bestimmten Klasse zugeordnet sind. Die Ränder der Gebiete werden als Entscheidungsgrenzen bezeichnet. Jeder Featurevektor x wird damit entsprechend seiner Lage eindeutig

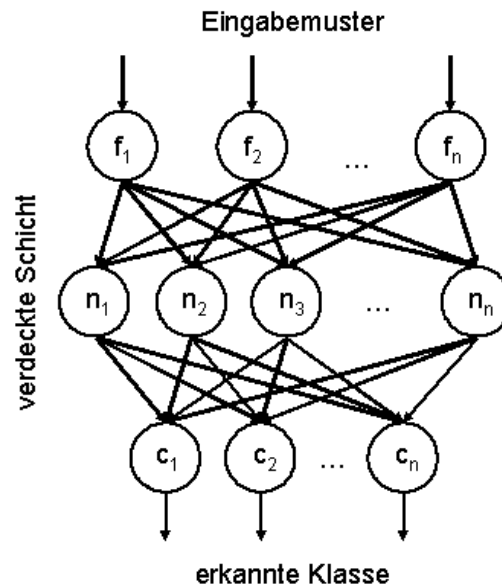


Abbildung 6.15: Der schematische Aufbau eines mehrschichtigen Perzeptrons.

einer Klasse c zugeordnet. Ein Klassifikator kann durch eine Menge von Diskriminanzfunktionen $g_1 \dots g_C$ repräsentiert werden (Abb. 6.16). Ein Muster x wird genau dann der Klasse c_i zugeordnet, wenn gilt:

$$g_i(x) > g_j(x) \forall j \neq i. \quad (6.19)$$

Lineare und quadratische Diskriminanzfunktion

Die lineare und quadratische Diskriminanzfunktion (LDF/QDF) zählt zu den statistischen Verfahren. Ausgehend von einem Trainingsset gelabelter Beispiele erstellt man ein Wahrscheinlichkeitsmodell für den Feature-Raum und ersetzt die Diskriminanzfunktion $g_i(x)$ durch die Wahrscheinlichkeit, dass x der Klasse c angehört [48]. Nach dem Bayes-Theorem errechnet sich die bedingte Wahrscheinlichkeit $P(B | A)$ für ein Ereignis B unter der Bedingung, dass A auftritt, zu:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}. \quad (6.20)$$

Angewandt auf ein Klassifizierungsproblem, bei dem die Klasse c einem Muster X zugeordnet werden soll, berechnet man zunächst für alle Klassen c_i die Wahrscheinlichkeit $P(c_i | x)$ und wählt dann die Klasse mit maximalem $P(c_i | x)$. Sei

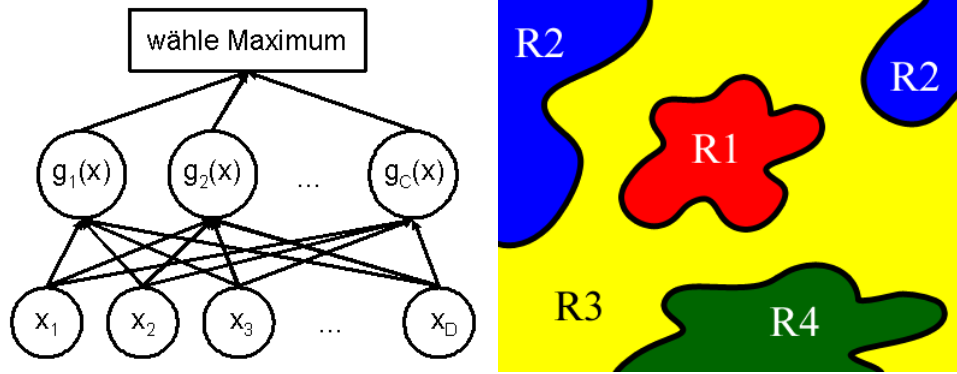


Abbildung 6.16: Repräsentation des Klassifikators durch eine Menge von Diskriminanzfunktionen.

$P(c_i)$ die A-Priori-Wahrscheinlichkeit für die Klasse c_i und $P(x | c_i)$ die Dichte der Verteilung der bekannten Muster für jede Klasse, so gilt unter der Annahme, dass $P(x)$ konstant ist:

$$P(c_i | x) = P(x | c_i)P(c_i) = \log(P(x | c_i)) + \log(P(c_i)). \quad (6.21)$$

Das Wahrscheinlichkeitsmaß $N(\mu, \Sigma)$ in \mathbb{R}^n ist durch die Standardnormalverteilung wie folgt definiert:

$$P(x | c_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right]. \quad (6.22)$$

Dabei bezeichnet μ_i den Mittelwertvektor und Σ_i die Kovarianzmatrix der Klasse c_i . Unter Anwendung des Bayes-Theorem erhält man also:

$$P(c_i | x) = \log \left(\frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right] \right) + \log(P(c_i)) \quad (6.23)$$

und schließlich nach Eliminierung konstanter Terme die folgende Formel, die als quadratische Diskrimanzfunktion bezeichnet wird:

$$g_i(x) = P(c_i | x) = -\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) - \frac{1}{2} \log(|\Sigma_i|) + \log(P(c_i)). \quad (6.24)$$

Der erste Term der quadratischen Diskrimanzfunktion entspricht der sog. Mahalanobis-Distanz, die in der statistischen Mustererkennung häufig als Distanzfunktion verwendet wird. Ersetzt man in 6.24 μ_i und Σ_i durch ihr globales Mittel, erhält man die lineare Diskrimanzfunktion:

$$g_i(x) = P(c_i | x) = -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) - \frac{1}{2} \log(|\Sigma|) + \log(P(c_i)). \quad (6.25)$$

Im Gegensatz zur quadratischen Diskriminanzfunktion sind die Grenzen der Entscheidungsbereiche nicht mehr quadratische Flächen sondern linear. Abbildung 6.17 zeigt die Normalverteilung und Entscheidungsgrenzen für die vier, in dieser Arbeit untersuchten Klassen.

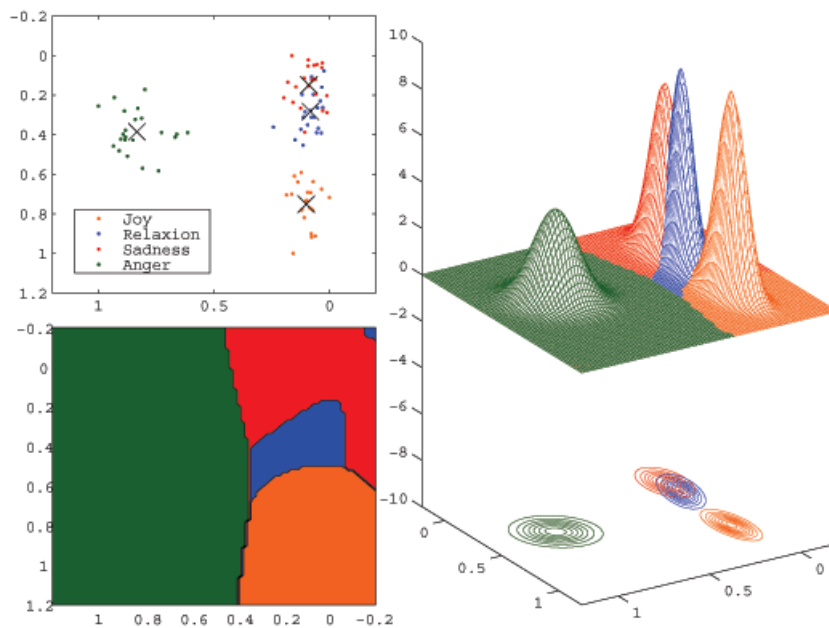


Abbildung 6.17: Die Figur oben links zeigt die Verteilung der Klassen nach Reduktion auf zwei Features mit der Fisher-Transformation. Auf der rechten Seite ist das Wahrscheinlichkeitsmodell dargestellt, das durch die zugehörige Standardnormalverteilung gegeben ist. Das Bild links unten zeigt die Entscheidungsregionen.

K-nächste Nachbarn

K -nächste Nachbarn (kNN) ist ebenfalls ein statistisches Verfahren, wird jedoch zu den instanzbasierten Verfahren gezählt, da es keine explizite Hypothese aufstellt, sondern die Trainingsbeispiele speichert und in der Klassifikationsphase mit den zu klassifizierenden Mustern vergleicht. Als Abstandsmaß wird dabei die euklidische Distanz verwendet:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6.26)$$

Im einfachsten Fall bekommt dann ein Muster die Klasse seines nächsten Nachbarn (1NN). Für $k > 1$, kann es jedoch vorkommen, dass die k -nächsten Nachbarn

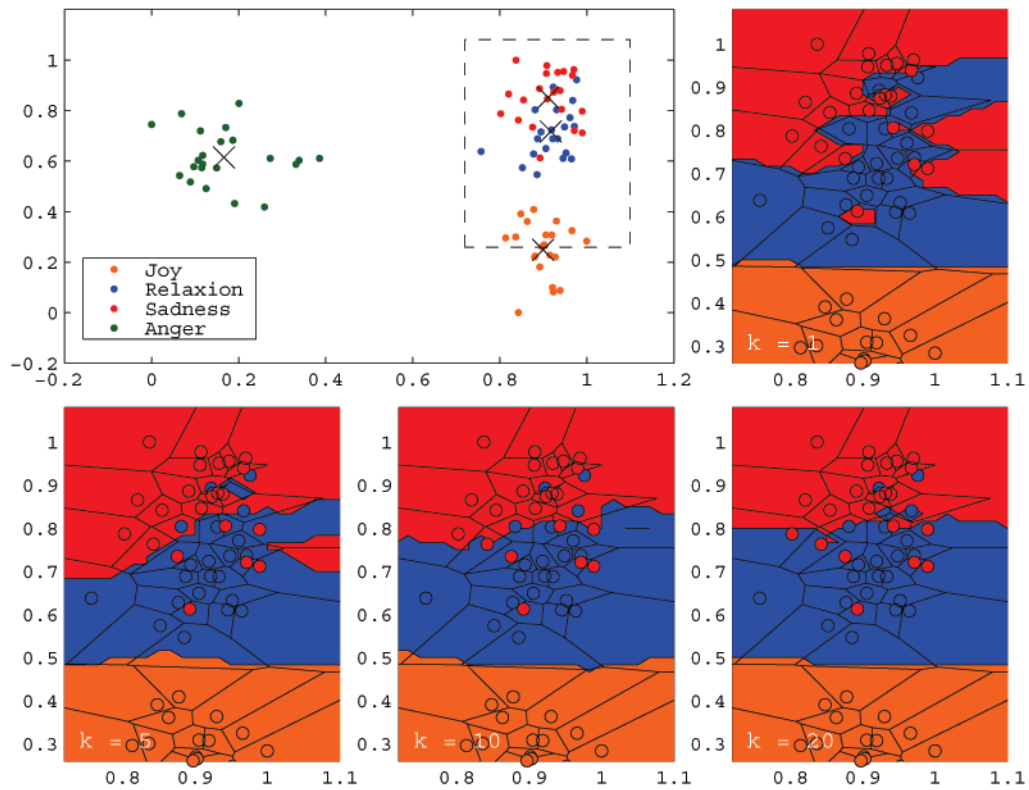


Abbildung 6.18: Die Anzahl der untersuchten Nachbarn hat insbesondere in Gebieten, in denen sich die Trainingsbeispiele verschiedener Klassen überlappen, Auswirkung auf den Verlauf der Entscheidungsgrenzen. Die Abbildung zeigt, wie sich mit wachsendem k die Grenzen glätten und „Inseln“, die durch Ausreißer bedingt werden, verschwinden. Die Daten wurden mit der Fisher-Transformation auf zwei Features reduziert, um eine grafische Darstellung zu ermöglichen.

verschiedene Label besitzen. In diesem Fall wird x die Klasse c_i zugeordnet, die in seiner Nähe am häufigsten auftritt. Sei $f(x, k)$ eine Funktion, die den k -nächsten Nachbarn zu x bestimmt und bezeichne $c_{f(x, k)}$ die zugehörige Klasse, dann gilt:

$$g_i(x) = \sum_{j=1}^k \delta(c_i, c_{f(x, j)}) \quad \text{mit } \delta(a, b) = \begin{cases} 1 & a = b \\ 0 & \text{sonst} \end{cases}. \quad (6.27)$$

Um den Algorithmus zu verbessern, kann zusätzlich eine Abstandsgewichtung eingeführt werden. Weiter entfernte Nachbarn werden dann weniger stark gewichtet

und haben damit einen geringeren Beitrag an der Klassifizierung:

$$g_i(x) = \sum_{j=1}^k \frac{\delta(c_i, c_{f(x, j)})}{d(x, f(x, j))}. \quad (6.28)$$

Um eine Division durch 0 zu verhindern, muss hier zunächst geprüft werden, ob x nicht bereits in den Trainingsbeispielen enthalten ist. Vorteile des Verfahrens sind, dass stets die gesamte Menge an Trainingsdaten unverändert verfügbar bleibt und sich leicht um neue Beispiele ergänzen lässt. Auch wenn die Anzahl der Trainingsinstanzen sehr groß ist, bleibt der Algorithmus effizient, da er stets nur die k -nächsten Nachbarn prüfen muss. Damit die Suche nach den Nachbarn nicht zu aufwändig wird, ist jedoch eine gute Indexierung nötig. Das Verfahren hat außerdem auch bei komplexen Funktionen noch Erfolg, allerdings sollte k nicht zu klein gewählt werden, um eine Approximation der Entscheidungsgrenzen zu erreichen und den Einfluss von Ausreißern zu verringern. Problematisch für das Verfahren wirkt sich die Verwendung der euklidischen Distanz als Abstandsmessung aus. Da jeweils alle Featurewerte von x berücksichtigt werden, können ähnliche Instanzen dennoch einen großen Abstand haben, wenn sie sich bereits in einem Feature stark unterscheiden. Da kNN instanzbasiert lernt, geschieht sämtliche Rechenarbeit erst während der Klassifizierung. Solche Verfahren werden deshalb auch als „faul“ bezeichnet. Abbildung 6.18 zeigt wie sich die Entscheidungsgrenzen bei gleichem Trainingsset aber verschiedenen Werten für k verändern.

Neuronale Netze

In dieser Arbeit wurde ein mehrstufiges Netz ohne Rückkopplung benutzt. Die Eingabeschicht besteht dabei aus gerade sovielen Zellen, dass sie ein kompletten Feature-Vektor aufnehmen kann. Diese Eingabezellen sind mit den Neuronen der verborgenen Schicht verbunden. Die Anzahl der verdeckten Einheiten wurde dabei zwischen zwei und acht variiert. Die dritte Schicht dient zur Ausgabe und enthält genau sovielen Neuronen wie untersuchte Klassen. Ein Feature-Vektor, der in das Netz eingespeist wird, aktiviert diese Zellen. Der Klassifikator entscheidet sich dann für die Klasse, deren Ausgabeneuron den größten Wert annimmt. Für die Implementierung des Netzes wurde die MATLAB-Toolbox „Netlab“ von Ian Nabney and Christopher Bishop verwendet [51]. Ein Beispiel für die Mustererkennung der physiologischen Daten mit dem beschriebenen neuronalen Netz ist in Abbildung 6.19 dargestellt.

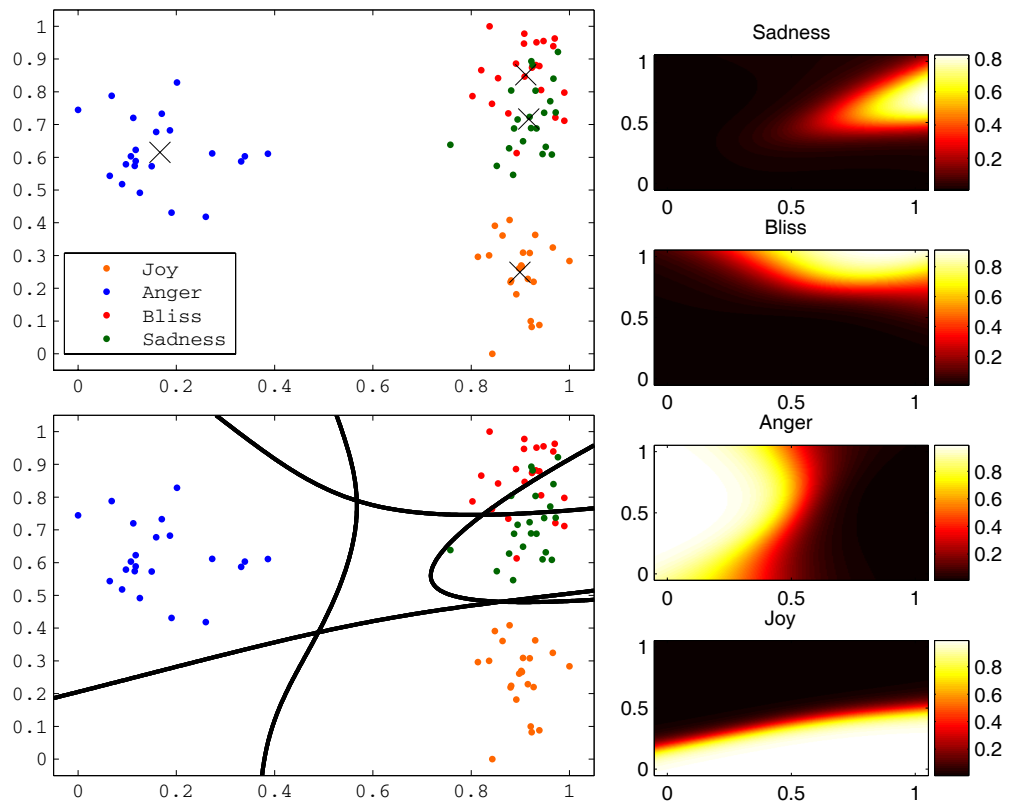


Abbildung 6.19: Die Figur oben links zeigt die Verteilung der Klassen nach Reduktion auf zwei Features mit der Fisher-Transformation. Auf der rechten Seite ist für jede Emotion die Netzausgabe nach 50 Iterationen dargestellt. Das Bild links unten zeigt schließlich den Verlauf der Entscheidungsgrenzen, die das Netz gelernt hat.

Kapitel 7

Ergebnisse

Es folgt eine Auflistung der Ergebnisse, die mit den in Kapitel 6 beschriebenen Methoden zur Dimensionsreduktion und Mustererkennung angewandt auf physiologische Daten, erzielt wurden. Getestet wurden die Verfahren sowohl mit eigenen Aufzeichnungen (vgl. Kapitel 5), als auch mit einem Datenkorpus des MIT [11]. Tabelle 7.1 gibt einen Überblick über Arbeiten, in denen ebenfalls statistische Verfahren oder neuronale Netze zur Klassifikation von physiologischen Daten eingesetzt wurden. Diese Arbeit testet Vertreter beider Methoden mit denselben Trainingsdaten, um einen direkten Vergleich der Ergebnisse zu ermöglichen. Weiter wurden statistische Analyseverfahren eingesetzt, um zu untersuchen wie sich die Werte eines Features in verschiedenen emotionalen Zuständen verändern. Dies ermöglichte es Zusammenhänge zwischen bestimmten Emotionen und Veränderungen in den Körpersignalen aufzudecken.

7.1 Test von Klassifikationsverfahren anhand des eigenen Korpus

Um die im letzten Kapitel beschriebenen Verfahren zur Mustererkennung zu testen, wurde ein eigener Korpus mit physiologischen Daten aufgezeichnet. Eine Beschreibung des Korpus wurde bereits in Kapitel 5 gegeben. Der folgende Abschnitt behandelt die Ergebnisse, die mit dem eigenem Korpus erzielt wurden.

7.1.1 Vier Emotionen ohne Dimensionsreduktion

Tabelle 7.2 zeigt die Ergebnisse, die für vier Emotionen mit 32 Features und ohne Dimensionsreduktion erzielt wurden. Die Erfolgsquote der Klassifikation wurde dabei mit der sog. „leave-one-out“-Methode bestimmt. Dabei wird jeweils ein Muster aus den verfügbaren Datensätzen herausgenommen. Mit den verbleibenden wird

Autor	Datenkorpus	Feature- zahl	Klassifi- kator	Reduk- tion	Ergebnisse
Healey [19]	8 Emotionen mit Cly- nes Methode [21]	11	DFA QDF	Fisher	80% - 90% für verschie- dene Emotionsgruppen
Vyzas [20]	8 Emotionen mit Cly- nes Methode [21]	40	DFA QDF	SFFS Fisher	81,25% für alle 8 Emo- tionen
Haag, Goronzy, Schaich, Williams [27]	je 3 positive und nega- tive Zustände mit va- riierendem Erregungs- level mit IAPS [28][29]	13	MLP	keine	Erregung mit 96,58% Lust mit 89,93%
Nasoz, Alvarez, Lisetti, Finkel- stein [30]	6 Emotionen mit Kurz- filmen und Filmaus- schnitten	k.A.	KNN DFA MBG ¹	keine	KNN: 71,6% DFA: 74,3% MBG: 83,7% jeweils für alle 6 Emo- tionen

¹ Marquardt Backpropagation

Tabelle 7.1: Übersicht über die Ergebnisse verwandter Arbeiten.

der Klassifikator aufgebaut. Danach wird das zuvor separierte Sample getestet. Der Reihe nach wird so jedes Muster genau einmal klassifiziert. Die Klassifikationsfehler ergibt sich dann aus der Anzahl der nicht richtig erkannten Muster durch die Gesamtzahl aller Beispiele. Der kNN-Klassifikator wurde mit $k = 3, 5, 10, 15$ getestet (3NN, 5NN, 10NN, 15NN), die Erkennungsquote blieb jedoch für alle vier Werte von k nahezu gleich und schnitt mit 79,55% bzw. 80,68% im Schnitt gleich oder leicht besser ab als LDF mit 79,55%. Mit 4 und 6 verdeckten Einheiten (MLP4 und MLP6) lieferte der MLP-Klassifikator ebenfalls 80,68%. Für 2 und 8 verdeckte Einheiten (MLP2 und MLP8) lag die Erkennungsrate nur noch bei 78,41%. Im Durchschnitt wurden die vier Emotionen unabhängig vom Klassifikator mit etwa 80% Wahrscheinlichkeit richtig zugeordnet. Unterschiede in der Erkennungsquote ergaben sich jedoch bei den einzelnen Emotionen. Während sie bei Ärger fast immer bei 100% lag, wurde Freude von 15NN nur mit 72,73%, von MLP4 und MLP6 dagegen mit 95,46% erkannt. Besondere Schwierigkeiten ergaben sich bei der Trennung von Glückseligkeit und Trauer. Im Durchschnitt wurden 20-30% der Beispiele jeweils der anderen Emotion zugeordnet. In den Tabellen 8.1-8.9 im Anhang findet sich noch einmal eine ausführliche Auflistung der Klassifikatoren und der Ergebnisse für jede der vier Emotionen.

Klassifikator	Freude	Ärger	Glückseligkeit	Trauer	Quote
LDF	77,27%	100%	72,73%	68,18%	79,55%
3NN	90,91%	100%	72,73%	59,09%	80,68%
5NN	86,36%	100%	72,73%	59,09%	79,55%
10NN	81,82%	100%	77,27%	63,64%	80,68%
15NN	72,73%	100%	77,27%	68,18%	79,55%
MLP2	90,91%	90,91%	72,73%	59,09%	78,41%
MLP4	95,46%	95,46%	63,64%	68,18%	80,68%
MLP6	95,46%	100%	63,64%	63,64%	80,68%
MLP8	86,36%	100%	59,09%	68,18%	78,41%

Tabelle 7.2: Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen ohne Dimensionsreduktion.

7.1.2 Vier Emotionen mit Dimensionsreduktion

Insgesamt vier Verfahren zur Dimensionsreduktion wurden getestet: einseitige sequentielle Suche (SFS), Fisher-Transformation (Fisher), eine Kombination aus SFS und Fisher (SFS/Fisher) und Varianzanalyse (ANOVA). Bei der hybriden Methode SFS/Fisher wurde zunächst eine Feature-Selektion mit SFS durchgeführt und anschließend die verbleibenden Features mit Fisher nochmals in einen kleineren Raum projiziert. Tabelle 7.3 zeigt die Ergebnisse der Dimensionsreduktion für die verschiedenen Verfahren. Mit SFS konnte für LDF eine Steigerung von 79,55% auf 92,05% erzielt werden. Mit SFS/Fisher in Kombination mit 5NN verbesserte sich die Klassifikationsrate von 79,55% auf 90,91%. In beiden Fällen wurde im SFS-Algorithmus als Evaluationsfunktion der Klassifikator selbst verwendet. Dies ist möglich, da beide Verfahren relativ schnell zu einem Ergebnis kommen. Ein neuronales Netz zu trainieren braucht dagegen sehr viel mehr Zeit. Aus diesem Grund wurde beim MLP-Klassifikator auf die Distanzfunktion von Wilk (vgl. 6.2.2) zur Evaluierung zurückgegriffen. In Kombination mit SFS/Fisher konnte für MLP6 die Klassifikationsquote von 82,82% auf 88,64% gesteigert werden. Mit Fisher allein konnte bei keinem der Klassifikatoren eine Verbesserung vermerkt werden. Mit ANOVA in Kombination mit LDF und 5NN immerhin noch 7,95% und 7,08%. Falls Performance eine Rolle spielt, bietet ANOVA also durchaus eine interessante Alternative zu der kostenaufwändigeren Suche mit dem SFS-Algorithmus. In Tabelle 8.10 im Anhang findet sich eine ausführliche Auflistung der Erkennungsquoten mit Dimensionsreduktion in Abhängigkeit der selektierten Features.

Klassifikator	keine	SFS	Fisher	SFS/Fisher	ANOVA
LDF	79,55%	92,05%	79,55%	90,91%	87,50%
5NN	79,55%	86,36%	80,68%	90,91%	86,36%
MLP6	81,82%	87,50%	80,68%	88,64%	86,36%

Tabelle 7.3: Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen mit Dimensionsreduktion.

7.1.3 Emotionsgruppen mit Dimensionsreduktion

Abhängig von der jeweiligen Anwendung ist man nicht immer daran interessiert bestimmte Emotionen zu erkennen, sondern möchte vielleicht lediglich wissen wann der Benutzer besonders erregt ist oder ob er sich in einem positiven oder negativen Zustand befindet. Tabelle 7.4 gibt einen Überblick über die Ergebnisse, die für drei verschiedene Emotionsgruppen erzielt wurden. Für negative Valenz wurden Ärger und Trauer, für positive Freude und Glückseligkeit zusammengefasst. Als niedrig erregend eingestuft wurden Trauer und Glückseligkeit, als stark erregend Ärger und Freude. Schließlich wurden noch die friedlichen Emotionen Trauer, Glückseligkeit und Freude mit Ärger verglichen. Niedrige und starke Erregung konnte mit etwa 95% von allen drei Klassifikatoren sehr gut getrennt werden. Erstaunlicherweise liegt die Erkennungsquote für negative und positive Zustände mit etwa 87% sogar unter der für alle vier Emotionen, obwohl statt vier nur noch zwei Klassen unterschieden werden müssen. Der Grund dafür ist wohl, dass durch die Kombination niedrig und hoch erregender Klassen inhomogene Klassen mit einer ungünstigen Verteilung der Feature-Vektoren entstehen.

Klassifikator	4 Emotionen	negativ/positiv	schwache/starke Erregung	friedlich/aufgebracht
LDF (SFS)	92,05%	86,36%	96,59%	100%
5NN (SFS/Fisher)	90,91%	86,36%	94,32%	100%
MLP6 (SFS/Fisher)	88,64%	88,64 %	94,32%	100%

Tabelle 7.4: Vergleich der Erkennungsquoten für verschiedene Klassifikatoren und Emotionsgruppen mit Dimensionsreduktion.

7.2 Bewertung der Features

Der Einsatz von Verfahren zur Feature-Selektion hat gezeigt, dass sich die Ergebnisse der drei Klassifikatoren deutlich verbessern lassen. Dabei fällt auf, dass meist eine kleine Anzahl von Features reicht, um die beste Klassifikationsraten zu

erzielen. Mit nur drei Features lassen sich z. B. die vier Emotionen von allen drei Klassifikatoren schon mit einer durchschnittlichen Wahrscheinlichkeit von über 80% erkennen. Ein Subset von 9 Features liefert bereits eine Erkennungsrate von über 90% (vgl. Tabelle 8.10 im Anhang) und nimmt für mehr Kennwerte wieder ab. Dies lässt vermuten, dass es für jede Emotion aus dem vorliegenden Set nur einige wenige, besonders signifikante Features gibt. In Kapitel 6.2 wurde bereits die Varianzanalysetechnik ANOVA zur Feature-Selektion eingeführt. Mit dem multiplen Vergleich von Mittelwerten nach Tukey können Features zwischen mehreren Klassen paarweise auf Signifikanz geprüft werden [52][34]. Tabelle 7.5 zeigt für jede Emotion die Features, die entsprechend dem Test von Tukey als signifikant eingestuft wurden. Ein + bzw. - hinter dem Namen des Kennwerts deuten an, dass

Signal	Freude	Ärger	Glückseligkeit	Trauer
SC	scMean + scStd + scDiff +	scMean + scDiff -	scMean - scStd -	scMean - scStd -
EMG	emgMean + emgStd +	emgMean + emgStd +	emgMean - emgStd -	emgMean - emgStd -
RSP	rspFreqMean - rspFreqStd - rspSpec3 + rspSpec4 - rspAmplMean +	rspFreqMean + rspFreqDiff + rspSpec1 + rspSpec2 - rspSpec3 - rspSpec4 + rspAmplMean -	rspFreqStd + rspFreqDiff - rspSpec2 + rspSpec3 - rspSpec4 -	rspSpec3 + rspSpec4 -
EKG	ekgFreqMean + ekgFreqStd + ekgFreqDiff +		ekgFreqMean + ekgFreqStd + ekgFreqDiff +	

Tabelle 7.5: Die Tabelle zeigt für jede Emotion die Features, die entsprechend dem Test von Tukey bei einem Signifikanzlevel von 0,05, als signifikant eingestuft wurden.

die Werte des Features im Mittel signifikant höher bzw. niedriger als der Durchschnitt sind. Dabei wurden jeweils die Muster einer Emotion gegen alle übrigen Muster getestet. Es zeigte sich, dass Freude durch ein hohes SC- und EMG-Level geprägt ist. Die Atmung wird tiefer und die Frequenz sinkt. Gleichzeitig steigt die Herzfrequenz. Ärger bedingt ebenfalls einen Anstieg des SC- und EMG-Niveaus. Die Atmung wird jedoch flacher und die Frequenz steigt an. Bei Glückseligkeit und Trauer ist das SC- und EMG-Level dagegen niedriger. Die Atemfrequenz ist bei Trauer leicht höher; bei Glückseligkeit unregelmäßiger. Wie bei Freude, wird

Glückseligkeit außerdem noch durch einen schnelleren Herzschlag begleitet. Es fällt auf, dass sich die Veränderungen in den Körpersignalen für Trauer und Glückseligkeit sehr ähneln. Hier liegt der Grund dafür, dass sich die beiden Emotionen nur schwer voneinander trennen lassen. Die gute Erkennungsrate für Ärger lässt sich dagegen auf die signifikant höhere Atemfrequenz zurückführen. Natürlich dürfen diese Ergebnisse nicht ohne weiteres verallgemeinert werden, da sie nur aus den Beobachtungen der Daten eines Individuums hervorgehen.

7.3 Vergleich mit dem Korpus des MIT

Der MIT-Korpus besteht aus insgesamt acht Emotionen, die in Tabelle 7.7 aufgelistet sind und enthält Signale von SC, EMG, RSP und BVP (Blood volume pulse). Aus letzterem kann wie beim EKG die Herzfrequenz bestimmt werden, so dass dieselben 32 Features wie für den eigenen Korpus berechnet werden konnten. Das SC-Signal wurde ebenfalls an der Handfläche abgenommen, die EMG-Elektroden jedoch an der Backe und nicht im Nacken angebracht. Daten von insgesamt acht

Korpus	Erregung	Emotionen	Signale	Größe
MIT	Bilder	keine Emotion Ärger Hass Kummer platonische Liebe romantische Liebe Freude Ehrfurcht	SC EMG RSP BVP	20 Tage 2 Minuten je Emotion
AuDB	Musik	Freude Ärger Glückseligkeit Trauer	SC EMG RSP EKG	22 Tage 2 Minuten je Emotion

Tabelle 7.6: Vergleich zwischen MIT- und AuDB-Korpus.

Emotionen enthält der MIT-Korpus. In Tabelle 7.7 sind die Emotionen mit ihrem jeweiligen Erregungs- und Valenzgrad aufgeführt. Um beide Korpora vergleichen zu können, wurden jeweils die schwach erregenden (keine Emotion, Hass, Kummer, platonische Liebe, Ehrfurcht) gegen die hoch erregenden (Ärger, romantische Liebe, Freude) bzw. die negativen (Ärger, Hass, Kummer) gegen die positiven Emotionen (platonische Liebe, romantische Liebe, Freude) getestet. Dies entspricht der Einteilung, die auch Healey [19] in ihren Untersuchungen verwendet. Für die jeweiligen

Emotion	Erregung	Valenz
keine Emotion	gering	neutral
Ärger	sehr hoch	sehr negativ
Hass	gering	negativ
Kummer	hoch	negativ
platonische Liebe	gering	positiv
romantische Liebe	sehr hoch	positiv
Freude	mittel hoch	positiv
Ehrfurcht	sehr gering	neutral

Tabelle 7.7: Überblick über die Emotionen aus dem Datenkorpus des MIT (vgl. Tabelle 3.1 in [19]).

Emotionsgruppen wurde dann die Klassifikationsquote bestimmt. Die Ergebnisse sind in in Tabelle 7.8 zusammengefasst. Schwache und starke Erregung wurde von LDF mit 87,50% und positive und negative Valenz mit 85% Wahrscheinlichkeit richtig erkannt. Die Tendenz, dass sich Emotionen in ihrem Erregungsgehalt besser unterscheiden lassen als in ihrem Lustanteil, bestätigte sich auch hier. Eine

Gruppe	positive/negative Valenz		schwache/starke Erregung	
	MIT	AuDB	MIT	AuDB
LDF (SFS)	85,00%	86,36%	87,50%	96,59%
5NN (SFS/Fisher)	81,67%	86,36%	86,88%	94,32%
MLP6 (SFS/Fisher)	82,50%	88,64%	86,25%	94,32%

Tabelle 7.8: Vergleich der Erkennungsquoten für verschiedene Klassifikatoren und Emotionsgruppen mit Dimensionsreduktion.

interessante Frage ist nun, inwiefern die Feature-Werte für beide Korpora und gleiche Emotionsgruppen auch ähnliches Verhalten zeigen, d. h. sich Zusammenhänge aus den physiologischen Daten beider Versuchspersonen ableiten lassen, oder ob diese doch zu individuell geprägt sind, um allgemeine Regeln zu finden. Dafür wurde wiederum die Signifikanz aller Features mit dem Tukey-Test geprüft. Tabelle 7.9 führt eine Übersicht über alle, als signifikant eingestuft Features. Sowohl im MIT-Korpus, als auch in den eigenen Aufzeichnungen, unterscheiden sich die Features mit starkem Erregungsgrad von solchen mit schwachen durch ein höheres SC- und EMG-Level und einen schnelleren Herzschlag. Aus dem Atemsignal lassen sich dagegen keine Gemeinsamkeiten ableiten. Während im MIT-Korpus eine hö-

Gruppe	starke vs. schwache Erregung		positive vs. negative Valenz	
Korpus	MIT	AuDB	MIT	AuDB
SC	scMean + scStd + scDiff +	scMean + scStd + scDiff -	scMean - scStd - scDiff +	scMean + scStd + scDiff +
EMG	emgMean + emgStd +	emgMean + emgStd +	emgMean - emgStd -	
RSP	rspFreqStd + rspFreqDiff - rspSpec1 + rspSpec2 + rspAmplMean + rspAmplStd +	rspFreqMean + rspFreqStd - rspFreqDiff + rspSpec2 - rspSpec4 +	rspFreqMean - rspSpec1 - rspSpec3 + rspAmplStd +	rspFreqMean - rspFreqDiff - rspSpec2 + rspSpec4 - rspAmplMean + rspAmplDiff +
EKG	ekgFreqMean + ekgFreqStd + ekgFreqDiff +	ekgFreqMean + ekgFreqStd + ekgFreqDiff +	ekgFreqMean -	

Tabelle 7.9: Die Tabelle zeigt für die Ergebnisse des Tukey-Signifikanztest für den MIT- und den eigenen Korpus bei einem Signifikanzlevel von 0,05 für hohe Erregung im Vergleich zu schwacher bzw. positive gegenüber negative Valenz.

here Amplitude und höhere Frequenzdichte im Bereich von 0-0.2 Hz als signifikant eingestuft werden, ist starke Erregung bei der Versuchsperson der AuDB durch eine höhere Atemfrequenz im Bereich von 0.3-0.4 Hz geprägt. Übereinstimmungen bei der Atmung finden sich dagegen bei Betrachtung von positiver und negativer Valenz. Sowohl im MIT-Korpus, als auch in der AuDB wird positive Valenz durch sinkende Atemfrequenz begleitet. Das SC-Niveau ist dagegen bei ersterem niedrig, bei der Versuchsperson der AuDB aber hoch. Andere Features sind jeweils nur für einen der beiden Probanden signifikant. Auch an dieser Stelle sei angemerkt, dass auch diese Betrachtungen eher als ein erster Versuch zu werten sind, mit statistischen Techniken und Methoden der Mustererkennung, allgemeine Zusammenhänge aus den Reaktionen der Körpersignale verschiedener Individuen im gleichen emotionalen Zustand abzuleiten. Aussagekräftige Ergebnisse können erst zukünftige Untersuchungen mit Daten von mehreren Versuchspersonen, die mit gleichem Versuchsaufbau und unter ähnlichen Bedingungen aufgezeichnet wurden, leisten.

Kapitel 8

Schluss

In der Arbeit wurden alle Schritte besprochen, die zur Entwicklung eines Systems notwendig sind, um anhand physiologischer Daten Rückschlüsse über den emotionalen Zustand eines Benutzers zu ziehen. Für jeden Schritt wurden jeweils verschiedene Methoden vorgestellt und implementiert. Um die Verfahren zu testen wurde ein eigener Datenkorpus aufgezeichnet, der auf einem Experiment basiert, bei dem Musikstücke zur Erregung der Emotionen benutzt werden. Verschiedene Überlegungen, die bei der Erstellung eines Korpus beachten werden sollten, wurden beschrieben. Der Datenkorpus enthält physiologische Daten zu den Emotionen Freude, Ärger, Trauer und Glückseligkeit. An 25 Aufzeichnungstagen wurden die Signale von den vier Sensoren EKG (Elektrokardiogramm), EMG (Elektromyogramm), RSP (Atmung) und SC (Hautleitwert) gesammelt. Diese wurden zunächst mit Techniken der digitalen Signalverarbeitung für den Mustererkennungsprozess aufbereitet und anschließend 32 Features berechnet. Verschiedene Klassifikatoren in Kombination mit unterschiedlichen Verfahren zur Dimensionsreduktion wurden getestet. So wurden sowohl statistische Verfahren, als auch neuronale Netze zur Klassifikation eingesetzt und die jeweiligen Ergebnisse gegenübergestellt. Mit Selektions- und Projektionsverfahren wurde anschließend untersucht, wie sich die Klassifikationsergebnisse in Kombination mit den verschiedenen Klassifikationsverfahren verbessern lassen. Alle vier Emotionen ließen sich mit den verwendeten Klassifikatoren gut unterscheiden. Mit den eingesetzten Verfahren zur Dimensionsreduktion konnten die Ergebnisse nochmal um über 10% verbessert werden. Mit Erfolgsquoten von über 90% für die vier Emotionen und über 95% für bestimmte Emotionsgruppen konnten Ergebnisse erzielt werden, die vergleichbar mit denen aus der Gesichts- und Spracherkennung sind. Mit einem speziellen statistischen Verfahren aus der Varianzanalyse wurde schließlich noch untersucht, wie sich die Emotionen mit bestimmten Körperreaktionen verknüpfen lassen. Ein, am MIT aufgezeichneter Korpus, wurde ebenfalls getestet und auf Gemeinsamkeiten und Unterschiede mit dem eigenen Korpus untersucht.

Anhang

LDF	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	77,27%	0%	9,09%	13,64%	77,27%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	72,73%	27,27%	72,73%
Trauer	9,09%	0%	22,73%	68,18%	68,18%
Quote	77,27%	100,00%	72,73%	68,18%	79,55%

Tabelle 8.1: Klassifikation von vier Emotionen mit LDF ohne Dimensionsreduktion.

3NN	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	90,91%	0%	9,09%	0%	90,91%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	72,73%	27,27%	72,73%
Trauer	9,09%	4,55%	27,27%	59,09%	59,09%
Quote	90,91%	100%	72,73%	59,09%	80,68%

Tabelle 8.2: Klassifikation von vier Emotionen mit 3NN ohne Dimensionsreduktion.

5NN	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	86,36%	0%	4,55%	9,09%	86,36%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	72,73%	27,27%	72,73%
Trauer	13,64%	4,55%	27,27%	59,09%	59,09%
Quote	86,36%	100%	72,73%	59,09%	79,55%

Tabelle 8.3: Klassifikation von vier Emotionen mit 5NN ohne Dimensionsreduktion.

10NN	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	81,82%	0%	4,55%	13,64%	81,82%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	77,27%	27,27%	77,27%
Trauer	9,09%	4,55%	22,73%	63,64%	63,64%
Quote	81,82%	100%	77,27%	63,64%	80,68%

Tabelle 8.4: Klassifikation von vier Emotionen mit 10NN ohne Dimensionsreduktion.

15NN	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	72,73%	0%	4,55%	22,73%	72,73%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	77,27%	22,73%	77,27%
Trauer	4,55%	4,55%	22,73%	68,18%	68,18%
Quote	72,73%	100%	77,27%	68,18%	79,55%

Tabelle 8.5: Klassifikation von vier Emotionen mit 15NN ohne Dimensionsreduktion.

MLP2	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	90,90%	4,55%	4,55%	0%	90,91%
Ärger	0%	90,91%	9,09%	0%	90,91%
Glückseligkeit	0%	4,55%	72,73%	22,73%	72,72%
Trauer	4,55%	4,55%	31,82%	59,09%	59,09%
Quote	90,91%	90,91%	72,73%	59,09%	78,41%

Tabelle 8.6: Klassifikation von vier Emotionen mit MLP2 ohne Dimensionsreduktion.

MLP4	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	95,46%	0%	0%	4,55%	95,46%
Ärger	0%	95,45%	0%	4,55%	100%
Glückseligkeit	4,55%	0%	63,64%	31,82%	63,64%
Trauer	9,09%	4,55%	18,18%	68,18%	68,18%
Quote	95,46%	95,46%	63,64%	68,18%	80,68%

Tabelle 8.7: Klassifikation von vier Emotionen mit MLP4 ohne Dimensionsreduktion.

MLP6	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	95,46%	0%	0%	4,54%	95,46%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	63,64%	36,36%	63,64%
Trauer	9,09%	4,55%	22,73%	63,64%	63,64%
Quote	95,46%	100%	63,64%	63,64%	80,68%

Tabelle 8.8: Klassifikation von vier Emotionen mit MLP6 ohne Dimensionsreduktion.

MLP8	Freude	Ärger	Glückseligkeit	Trauer	Quote
Freude	86,36%	0%	4,55%	9,09%	86,36%
Ärger	0%	100%	0%	0%	100%
Glückseligkeit	0%	0%	59,09%	40,91%	59,09%
Trauer	4,55%	4,55%	22,73%	68,18%	68,18%
Quote	86,36%	100%	59,09%	68,18%	78,41%

Tabelle 8.9: Klassifikation von vier Emotionen mit MLP8 ohne Dimensionsreduktion.

Dimension	LDF	5NN	MLP6
Reduktion	SFS	SFS/Fisher	SFS/Fisher
1	56,82%	56,82%	59,09%
2	79,55%	78,41%	79,55%
3	84,01%	85,23%	81,82%
4	86,36%	88,64%	78,41%
5	87,50%	87,50%	79,55%
6	89,71%	88,64%	81,82%
7	90,91%	87,50%	81,82%
8	92,05%	87,50%	84,10%
9	92,05%	88,64%	87,50%
10	92,05%	87,50%	87,50%
11	92,05%	88,64%	88,64%
12	90,91%	88,64%	88,64%
13	92,05%	87,50%	87,50%
14	92,05%	85,23%	85,23%
15	90,91%	85,23%	84,09%
16	89,77%	86,36%	85,23%
17	88,64%	89,77%	87,50%
18	87,50%	89,77%	87,50%
19	87,50%	90,09%	82,95%
20	87,50%	90,09%	82,95%
21	85,23%	90,09%	80,68%
22	87,50%	89,77%	78,41%
23	88,64%	87,50%	77,27%
24	87,50%	87,50%	76,14%
25	86,36%	86,36%	77,27%
26	85,23%	84,10%	80,68%
27	84,09%	84,10%	79,55%
28	82,95%	81,82%	79,55%
29	82,95%	81,82%	78,41%
30	82,95%	82,95%	77,27%
31	81,18%	81,82%	79,55%
32	79,55%	79,55%	78,41%

Tabelle 8.10: Vergleich der Erkennungsquoten für verschiedene Klassifikatoren bei vier Emotionen mit Dimensionsreduktion in Abhängigkeit der selektierten Features.

Literaturverzeichnis

- [1] P.R. Kleinginna, A.M. Kleinginna: A categorized list of emotion definitions with suggestions for a consensual definition, 1981
- [2] Prof. Dr. Joachim Funke: Onlinekurs - Allgemeine Psychologie,
<http://www.psychologie.uni-heidelberg.de/ae/allg/lehre/wct/>
- [3] A.M. Isen, B. Means, R. Patrick, G.P. Nowicki: Some factors influencing decision-making strategy and risk taking, 1982
- [4] A.M. Isen, K.A. Daubman, G.P. Nowicki: Positive affect facilitates creative problem solving, 1987
- [5] Paul Ekman: An argument for basic emotions. Cognition and Emotion, 1992
- [6] A. Ortony, T.J. and Turner: What's basic about basic emotions?, 1990
- [7] Wilhelm Wundt: Principles of Physiological Psychology (Volume 1, 2nd Edition), 1910
- [8] P.J. Lang, M.K. Greenwald, M.M. Bradley: Looking at pictures: Affective, facial, visceral and behavioral reactions, 1993
- [9] Lynellen D.S. Perry: Emotionware,
<http://www.acm.org/crossroads/xrds3-1/emotware.html>
- [10] Rosalind W. Picard: What does it mean for a computer to „have” emotions?, MIT Media Laboratory
- [11] MIT: Affective Computing,
<http://affect.media.mit.edu/>
- [12] Rosalind W. Picard, Jennifer A. Healey: Affective Wearables, 1997
- [13] Changbo Hu, Rogerio Feris, Matthew Turk: Real-time View-based Face Alignment Using Active Wavelet Networks, IEEE Workshop on Analysis and Modeling of Faces and Gestures, ICCV'03, Nice, France, 2003

- [14] Jonathan Klein, Rosalind W. Picard: Support for human emotional needs in human-computer interaction, Affective Computing Group
- [15] Eu Jin Teoh, Daniel Seet: PLASTIC,
www.scholars.nus.edu.sg/resources/iface/design/condiment/interface.html
- [16] Brenda Laurel: Computers as Theatre, 1993
- [17] Jennifer A. Healey, Rosalind W. Picard, Frank Dabek: A New Affect-Perceiving Interface and It's Application to Personalized Music Selection, MIT Media Laboratory
- [18] Jocelyn Scheirer: Galvactivator Website,
<http://www.media.mit.edu/galvactivator/index2.html>
- [19] Jennifer A. Healey: Wearable and Automotive Systems for Affect Recognition from Physiology, 2000
- [20] Elias Vyzas: Recognition of Emotional and Cognitive States Using Physiological Data, 1999
- [21] M. Clynes: Sentic: The Touch of the Emotions, 1977
- [22] Rosalind W. Picard: Affective Computing, MIT Press 1997
- [23] Rosalind W. Picard: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State
- [24] Thought Technology Ltd.: Webseite der Firma,
<http://www.thoughttechnology.com/>
- [25] Rainer Schandry: Lehrbuch Psychophysiologie, BELTZ Psychologie Verlags Union, 1981
- [26] Julien Fistre, Atau Tanaka: Real Time EMG Gesture Recognition for Consumer Electronics Device Control, 2002
- [27] Andreas Haag, Silke Goronzy, Peter Schaich, Jason Williams: Emotion Recognition Using Bio-Sensors: First Step Towards an Automatic System
- [28] P.J. Lang, M.M. Bradley, B.N. Cuthberth: International Affective Picture System (IAPS): Instruction Manual and Affective Ratings, University of Florida, 2001
- [29] Center for the Study of Emotion and Attention: The International Affective Picture System: Digitized Photographs, University of Florida, 2001

- [30] Fatma Nasoz, Kaye Alvarez, Christine L. Lisetti, Neal Finkelstein: Emotion Recognition from Physiological Signals for Presence Technologies, 2003
- [31] Fatma Nasoz, Kaye Alvarez, Christine L. Lisetti, Neal Finkelstein: Emotion Recognition from Physiological Signals for User Modeling of Affect
- [32] The MathWorks - MATLAB and Simulink for Technical Computing, <http://www.mathworks.com/>
- [33] Anil K. Jain, Robert P.W. Duin, Jianchang Mao: Statistical Pattern Recognition: A Review, 2000
- [34] NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>
- [35] B. Weaver, One-way ANOVA, 2003
- [36] K. Fukunaga: Introduction to Statistical Pattern Recognition, 1990
- [37] Anil K. Jain, B. Chandrasekaran: Dimensionality and Sample Size Considerations in Pattern Recognition Practise, 1982
- [38] M. Dash, H. Liu: Feature Selection for Classification, 1997
- [39] P. Pudil, F.J. Ferri, J. Novovicova, J. Kittler: Floating Search Methods for Feature Selection with Nonmonotonic Criterion Functions
- [40] R.O. Duda, P.E. Hart: Pattern Classification and Scene Analysis, 1973
- [41] S. Balakrishnama, A. Ganapathiraju: Linear Discriminant Analysis - A Brief Tutorial
- [42] W. Zhao, R. Chellappa, N. Nandhakumar: Empirical Performance Analysis of Linear Discriminant Classifiers
- [43] ShiXin Yu: Feature Selection and Classifier Ensembles: A Study on Hyperspectral Remote Sensing Data, 2003
- [44] Ira Cohen, Qi Tian, Xiang Sean Zhou, Thomas S. Huang: Feature Selection Using Principal Feature Analysis
- [45] S. Bermejo, K. Cabestany: Oriented Principal Component Analysis for Feature Extraction, 1999
- [46] Robert T. Olszewski: Generalized Feature Extraction for Structural Pattern Recognition in TimeSeries Data, 2001

- [47] Romain Chabbal, Reconnaissance de l'état émotionnel d'un étudiant interagissant avec un jeu vidéo éducatif, 2002
- [48] B.D. Ripley: Linear and Quadratic Discriminant Analysis, 2003
- [49] W. Pitts, W.S. McCulloch WS: W. A logical calculus of the ideas immanent in nervous activity, 1943
- [50] R.M. Hristev, The ANN Book, 1998
- [51] Ian Nabney: Netlab: Algorithms for Pattern Recognition
- [52] Y. Hochberg, A. C. Tamhane, Multiple Comparison Procedures, 1987